HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

WHAT MAKES A GREAT DECATHLETE?

Helsingin yliopisto
Matemaattis-luonnontieteellinen tiede-
kunta
Matematiikan ja tilastotieteen osasto
Pro gradu tutkielma
Tilastotiede
Toukokuu 2020
Otto Ylöstalo
014748280

Ohjaaja: Petteri Piiroinen, Kimmo
Vehkalahti

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta - Fakultet – Faculty Matematisk-naturvetenskapliga fakulteten	Laitos - Institution – Department Institutionen för matematik och statistik	
Tekijä - Författare – Author Otto Ylöstalo		
Työn nimi - Arbetets titel Pro gradu avhandling		
Title What makes a great decathlete?		
Oppiaine - Läroämne – Subject Statistik		
Työn laji/ Ohjaaja - Arbetets art/Handledare – Level/Instructor Magisteravhandling / Petteri Piiroinen, Kimmo Vehkalahti	Aika - Datum - Month and year 18.5.2020	Sivumäärä - Sidoantal - Number of pages 39 pp.

Tiivistelmä - Referat – Abstract

Tiokamp är en friidrottsgren som består av tio enskilda grenar som utförs i en bestämd ordning under två dygn. Resultaten omvandlas enligt en officiell tabell till poäng vars summa resulterar i det slutgiltiga resultatet. Avhandlingens mål är att undersöka vad som gör en elitmångkampare. Den individuella utvecklingen i varje enskild gren utvärderas.

Som material har använts alla tiokampare som samlat över 8 500 poäng mellan åren 1970 och 2019 och alla deras mångkampsrelaterade prestationer fram till året då idrottaren gjort sina högsta poäng. Mångkampsprofilens utveckling undersöks i fyra nivåer. Möjliga skillnader i utvecklingen observeras utgående från den inledande nivån hos idrottaren.

I teorin behandlas Laird och Wares flernivåmodell (LME model) samt en närmare härledning av restricted maximul likelihood-metoden (REML-metoden) som presenteras i sambandet.

Resultaten i avhandlingen tyder på att sprintgrenarna spelar en stor roll i tiokampen, framförallt förutsätter 8 500 poäng en viss specifik standard på 400 meter. Det ter sig också att inledningsvis svaga idrottare i en specifik gren utvecklas stort under de första stegen i karriären.

Avainsanat – Nyckelord
friidrott, tiokamp, mångkampsprofil, regression, lineär modell

Keywords
athletics, track and field, decathlon, decathlete’s profile, multilevel model, hierarchical model, linear mixed-effects model

Säilytyspaikka - Förvaringsställe - Where deposited

Muita tietoja - Övriga uppgifter - Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta - Fakultet – Faculty	Laitos - Institution – Department	
Matemaattis-luonnontieteellinen	Matematiikan ja tilastotieteen laitos	
Tekijä - Författare – Author		
Otto Ylöstalo		
Työn nimi - Arbetets titel		
Pro gradu tutkielma		
Title		
What makes a great decathlete?		
Oppiaine - Läroämne – Subject		
Tilastotiede		
Työn laji/ Ohjaaja - Arbetets art/Handledare – Level/Instructor	Aika - Datum - Month and year	Sivumäärä - Sidoantal - Number of pages
Maisterin tutkielma / Petteri Piiroinen, Kimmo Vehkalahti	18.5.2020	39 pp.

Tiivistelmä - Referat – Abstract

Kymmenottelu on yleisurheilulaji, joka koostu kymmenestä yksilölajista, jotka suoritetaan samassa järjestyksessä kahden vuorokauden aikana. Kymmenen lajin tulokset pisteytetään virallisen pistetaulukon mukaan ja lopputulos on näiden lajien yhteenlaskettu pistemäärä. Tutkielman tavoitteena on tutkia, mikä tekee huippuluokan kymmenottelijaa. Lisäksi tarkastellaan miten henkilökohtaiset kehityskäyrät vertautuvat lajeittain.

Aineistona käytetään vuonna 1970-2019 yli 8 500 pisteen kymmenottelijoiden kaudenparhaiden ottelusarjassa tehtyjen tuloksien tietoja. Ottelijaprofiilin muutoksia tutkitaan neljässä tasossa. Erikseen tarkastellaan ottelijan henkilökohtainen kehitys lajeittain ja näiden mahdollisia eroja.

Teoriaosassa käsitellään Lairdin ja Waren lineaarinen sekamalli (LME malli) sekä siihen liittyvä rajoitettu suuremman uskottavuuden menetelmää (REML).

Tutkielman tuloksien valossa vaikuttaa siltä, että kymmenottelussa pikajuoksulajit ovat tärkeimmässä roolissa. Varsinkin lähtökohtaisesti heikoille 400 metrin juoksijoille kehitys on suuri. Kymmenottelijan profiilissa on havaittavissa, että tietyt lähtökohtaisesti heikot lajit pysyvät suhteellisen heikkona verrattuna muihin.

Avainsanat – Nyckelord
yleisurheilu, kymmenottelu, ottelijaprofiili, lineaariset sekamallit, monitasomallit, hierarkkiset mallit

Keywords
athletics, track and field, decathlon, decathlete’s profile, linear mixed-effects model, multilevel model, hierarchical model

Säilytyspaikka - Förvaringsställe - Where deposited

Muita tietoja - Övriga uppgifter - Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta - Fakultet - Faculty Faculty of Science		Laitos - Institution – Department Department of Mathematics and Statistics
Tekijä - Författare – Author Otto Ylöstalo		
Työn nimi - Arbetets title Pro gradu thesis		
Title What makes a great decathlete?		
Oppiaine - Läroämne – Subject Statistics		
Työn laji/ Ohjaaja - Arbetets art/Handledare – Level/Instructor Master’s Thesis / Petteri Piiroinen, Kimmo Vehkalahti	Aika - Datum - Month and year 18.5.2020	Sivumäärä - Sidoantal - Number of pages 39 pp.

Tiivistelmä - Referat – Abstract

The decathlon is a track and field event that consists of ten single events, which are performed in the same order during two competitive days. The result from each event is according to the official scoring table transformed into points which are then added together and result in the final score. The aim of this study is to see, what makes an elite decathlete. The individual development in each event is observed.

For this study material consisting of all recorded data by decathletes with a personal best of over 8 500 points from year 1970 to 2019 have been used. From the data the best (individual) results made within a decathlon per year was chosen up until the best season. Changes in the decathlete´s profile has been analyzed in four stages (clusters). Changes in the individual development pattern based on the initial level in each event is also observed.

The theory part covers the linear mixed-effects model introduced by Laird and Ware. The restricted maximum likelihood method related to the model is deduced in detail.

According to the results of the study it seems like the sprint events are of great importance in the decathlon. Especially initially weak 400-meter-runners make a notable progress. However, some initially weak events remain relatively weak compared to the initially stronger.

Avainsanat – Nyckelord

Keywords
athletics, track and field, decathlon, decathlete’s profile, linear mixed-effects model, multilevel model, hierarchical model

Säilytyspaikka - Förvaringsställe - Where deposited

Muita tietoja - Övriga uppgifter - Additional information

CONTENT

1 INTRODUCTION

1.1 The Decathlon

The decathlon is a track and field event that consists of ten single events: four track events and six field events. The result from each event is according to the official scoring table (IAAF, 2001) transformed into points which are then added together resulting in the final score. The events are always performed in the same order during two competitive days: 100m, long jump, shot put, high jump, 400m, 110m hurdles, discus throw, pole vault, javelin throw and 1500m. The first day ends with the 400m run and the decathlon finishes with the 1500m run the following day, hence dividing the ten events into five events per day (IAAF, 2001). Officially the decathlon is only an event for men, the women's decathlon is not an official event and they compete in the heptathlon (Zarnowski, 1989). In this study the events are often referred to as groups where the sprints include the 100m, 400m and 100m hurdles and the jumps the long jump, high jump and pole vault. The throws are represented by the shot put, discus throw and javelin throw while the 1 500m run is not considered belonging to any group.

1.2 History

Based on the averages of the world's 100 best decathletes (by 2005) through history events such as 100m, long jump and 110m hurdles are more favorable for decathletes compared to the throwing events and the 1500m (Westera, 2006). When looking at the participants in the Olympic Games 1938, 1948 and 1952 the 110m hurdles correlate the most to the final score while the pole vault, javelin throw and the 1 500m were the most unfavorable (Karvonen & Niemi, 1953). There are four standard types of decathletes based on which events they per percent gain the most points. There is the sprinter-jumper, the sprinter-thrower, the jumper-thrower and the universal type. All these types are well represented among today's decathletes, with a small majority of sprinter-jumper and universal types (Bauersfeld & Schröter, 2015. 661-662.).

The choice of topic relates to the author´s career as a national level decathlete left curious what only took him to the door of the international stage. In this study the author will continue his unpublished Bachelor´s thesis where he studied the differences between decathletes on different levels and what feature the considered milestone of 8

000 points demands. The earlier study suggests that being on a certain level in the sprint-based events is one common feature among good decathletes and that the maintenance of a high level in those events seems important while developing the throws and jumps (Ylöstalo, 2018). This study returns to the theme and tries to answer these questions on a broader and more individual basis.

1.3 Aim of the study

The aim of this study is to briefly compare decathletes that have scored over 8 500 points and to make a hypothetical pattern of development for these athletes. This could be of great value to understand taking the whole sport to another level. When it comes to all the events in the decathlon it could be important for coaches and athletes to know which events affect each other the most and if there is any feature of greater importance. It would be beneficial when planning the training program to realize which events needs the most attention. To answer this question statistical methods will be used featuring linear mixed-effects modeling. Individual linear models will be designed and studied for each event based on the level of the athlete in the decathlon. In the model the athlete's peak performance in each event on each level will be obtained. This approach differs from most studies of the decathlon where you only look at the best fulfilled decathlons where athletes may have underperformed in some event and hence not giving reliable information about the athlete's true potential. One aim is to discover if this method is appliable on this kind of data.

The linear mixed-effects model introduced by Laird and Ware (1982) will be applied and the restricted maximum likelihood method related to the model will be covered in detail in Appendix A. The understanding of the linear mixed-effects model presumes basic understanding of linear regression modeling and Bayesian inference.

2 MATERIAL

2.1 Data description

The data contains all decathlons (totally 1 834) made by 61 decathletes who to this date have scored over 8500 points in the decathlon in a compact format (Table 2) representing four stages, here named levels, on their way to their personal best. The first athlete to score over 8500 points according to today's scoring system was in 1976 and the most recent in 2019. The oldest observations from earlier stages in the development are from 1970. There are totally 107 missing values in the series used. A total number of decathletes with missing stages (years) of their career are 21. All results have then been transformed to points according to the scoring table of the International Association of Athletics Federation (IAAF). Table 1 shows the robust version where each athlete´s (`id`) best individual performance within a decathlon each year (`c_y`) have been noticed. This approach is different from most studies because it doesn´t require the results being from the same decathlon. The events are in the competitive order with the highest decathlon points (`SCORE`) per year and the athlete´s personal best (`PB`) following. Results made with junior implements and missing results are represented by `NA`.

Table 1. The data after modifications.

```
> data
      id  c_y M100  LJP SPP HJP M400 MH110  DP PVP   JP M1500 SCORE   PB
  1:   1 1979  765  790 693 714  745   857 631 702  905   776  7387 8709
  2:   1 1980  769  778 693 749  763   878 665 731  989   767  7577 8709
  3:   1 1981  789  807 768 767  797   875 669 760  963   804  7962 8709
  4:   1 1982  821  891 806 776  831   927 718 790  945   806  8104 8709
  5:   1 1983  836  962 835 794  839   980 713 790  817   752  8270 8709
 ---
502:  60 1996  952 1079 731 840  893  1044 768 849  842   757  8706 8706
503:  61 2015  746  757  NA 850  772    NA  NA 819   NA   666  7200 8691
504:  61 2016  750  764  NA 896  829    NA  NA 849  914   800  7800 8691
505:  61 2018  789  883 741 887  889   878 794 819  870   842  8220 8691
506:  61 2019  823  876 801 831  886   894 854 910 1028   841  8691 8691
```

2.2 Collection and modification of data

For this study material has been gathered by hand from Decathlon 2000 consisting all recorded initiated decathlons by decathletes with a personal best over 8500 points to date. Decathlon 2000 is the world´s largest and comprehensive database for the decathlon and hence a reliable source for this study. From the data there are a total of 1 834 decathlons made by 61 decathletes starting from the year 1970. To better balance the data due to big variations in the number of decathlons each decathlete´s best recorded individual event within a decathlon per year has been selected to fully describe the athlete's potential during that year (Table 1). This means, that for some decathletes the decathlon score may not add up to the sum of the individual results that year.

To support this study the data has been reduced to include only scores leading up to the year when the decathlete scored his highest mark to date. Results made with junior implements (shot put, 110m hurdles, discus throw and javelin throw) have been removed.

Since every decathlete has his own development path and some have missed seasons due to various reasons one further modification has been made; the decathlete´s progress is divided into levels (scoring groups) instead of years. Four levels have been selected: under 7500 points for group 1, 7500-7999 points for group 2, 8000-8499 points for group 3 and over 8500 points for group 4. In Table 2 these are named "`score_group`".

A decathlete reaches a new level the first year he achieves a total score that falls within the marks noted. After a decathlete has reached a new level he is considered belonging to that level until he takes the next step regardless of years with weaker total score in between. An athlete can´t move down from a level.

Since an athlete can stay on one level for more than one year it means that the best scores are not necessarily even from the same year. For example, in Table 2 the highest score in the 100 meters during each level is named "`M100_max`". The number of years spent on a level (`N`) is not relevant for the analysis.

In each level the best individual score from each event has been selected to represent the decathlete´s potential in each phase. Years that do not include any recorded decathlons are excluded. These do not affect the athlete´s achieved level. Some uncompleted decathlons are exceptionally included if they have results from at least eight (8) events. Table 2 shows the data after these modifications that include a total of 223 observations where 21 decathletes are missing one or two of the stages.

The level (`score_group`) will be our covariate in our model while the score for each event at each level will be the response. All single events will be analyzed separately. This is also the final modification made on the data before applying the linear mixed-effect model.

Table 2. Decathletes per scoring group. Years without decathlon scores are not included.

```
> prog
     id score_group M100_max LJ_max SP_max HJ_max M400_max MH110_max D_max PV_max JT_max M1500_max SCORE N
  1:  1           1      765    790    693    714      745       857   631    702    905       776  7387 1
  2:  1           2      789    807    768    767      797       878   669    760    989       804  7962 2
  3:  1           3      836    962    835    794      839       980   718    790    945       806  8270 2
  4:  1           4      890    962    864    803      890       984   829    880    924       820  8709 1
  5:  2           1      755    859    699    859      671       845   667    790    688       594  7298 4
 ---
219: 60           4      952   1079    731    840      893      1044   768    849    842       757  8706 1
220: 61           1      746    757     NA    850      772        NA    NA    819     NA       666  7200 1
221: 61           2      750    764     NA    896      829        NA    NA    849    914       800  7800 1
222: 61           3      789    883    741    887      889       878   794    819    870       842  8220 1
223: 61           4      823    876    801    831      886       894   854    910   1028       841  8691 1

> prog[id == 60]
   id score_group M100_max LJ_max SP_max HJ_max M400_max MH110_max D_max PV_max JT_max M1500_max SCORE N
1: 60           2      933    903    672    831      796       963   678    819    783       697  7938 1
2: 60           4      952   1079    731    840      893      1044   768    849    842       757  8706 1
```

As we can see from the data just by comparing decathlete id´s 1 and 61 is, that decathlete 1 needed more years to go beyond 8500 points because of spending more time a level 2 (`N = 2`) and level 3 (`N = 2`) while decathlete 61 managed to do this in four years. Since results made with junior implements have been excluded (`NA`) it also shows that id 61 was considered belonging to level 1 and level 2 during his junior years.

Decathlete id 60 is also shown as an example of a decathlete where levels 1 and 3 are missing because there are some individuals with exceptional career paths due to most commonly injuries affecting the development.

Table 3 shows an overview of these stages. The events are in the same order as in Table 1 and Table 2. Here the median score from Table 2 has been converted into what the given score would mean in result based on the IAAF scoring table. As an example, the median 100m score at level one was calculated to 808 points which means a time of 11.24 seconds etc. The total score is the sum of all these medians at each level.

Table 3. Median results by level.

Group	100M	LJ	SP	HJ	400M	110MH	DT	PV	JT	1500M	Total
1	11.24	712	13.04	200	50.24	15.09	39.95	430	56.58	4:36.65	7521
2	11.04	736	13.94	203	49.50	14.77	42.36	460	57.59	4:32.77	7956
3	10.80	760	15.10	207	48.34	14.27	46.54	490	64.42	4:26.83	8614
4	10.65	773	15.55	208	48.05	14.09	47.99	500	65.67	4:26.83	8832

Figure 1 demonstrates how the relative amount of points coming from the different event groups in the process changes. The 1500m is not included.

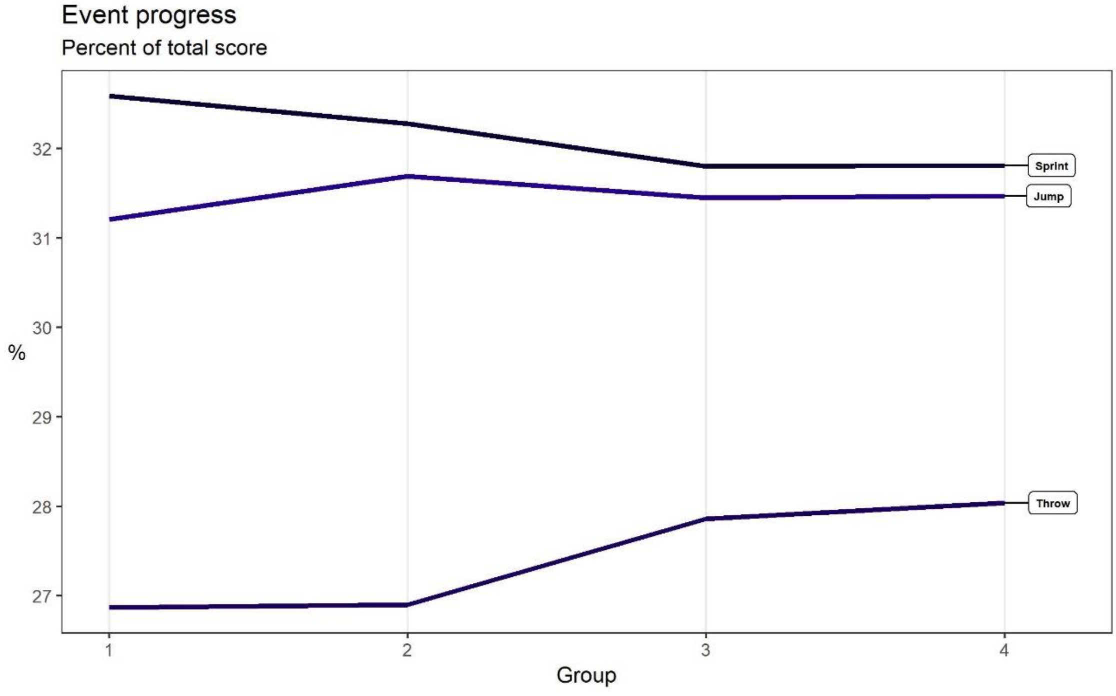


Figure 1. The percentage amount of points by median coming from the sprint-based events, jumps and throws.

3 METHODS

3.1 Linear models

For one to understand the meaning and benefits of linear mixed effect models some prior knowledge about basic linear modeling is required. Linear modeling is a common tool in economic and social science where researchers often use data collected over time or data in some other way categorical. The primary goal of these (longitudinal) studies is to identify changes over time and to discover factors that influence the changes.

A general model for linear modeling is

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{b} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ represents a N-dimensional vector of observations, $\mathbf{X}$ a $N \times (k+1)$ design matrix of the form $[\mathbf{1}_N\ \mathbf{X}_1]$ with the N-dimensional vector-column $\mathbf{1}_N$ of 1´s and the $N \times\ k$ matrix $\mathbf{X}_1$ containing independent variables x_i. The regression parameters to be estimated are represented by the $(k+1)$-dimensional vector $\boldsymbol{b}$ and the errors by the $N \times 1$ independent and identically distributed random variable vector $\boldsymbol{\varepsilon}$ (Gruber & Searle, 2017, 2).

The standard linear model relies on three basic assumptions (Agresti, 2015, 26):

1. All observations are normally distributed i.e. $y_i \sim N(\mu_i, \sigma_i^2)$
2. All observations have the common variance $\sigma_i^2 = \sigma^2$
3. The linear predictor and the mean have the exact relationship $x_i^\top \boldsymbol{b} = \mu_{\boldsymbol{i}}$ where $x_i^\top$ holds the covariate variables $(x_{i1}, \dots, x_{ik})$ and $\boldsymbol{b}$ is a coefficient vector

This can lead to challenges in the estimation of the error terms since the standard model does not consider grouping of observations. Linear least squares are a common way of estimation in linear modeling. This can cause inaccuracies in the error terms since they are assumed independent with equal error variances. Even though there are adapted methods to handle these issues they can still be fragile (Garson, 2013, 6).

In our case $\boldsymbol{y} \in \mathbb{R}^{223\times1}$ is a vector where y_i would mean the score in one event for the athlete with id $i = 1, \dots, 61$ and x_i the level of the athlete with id i. The vector $\boldsymbol{b}$ in our case is fitted by ordinary least squares (OLS) and $k = 1, \dots, 4$. Table 2 we hence obtain the variables for the 100 meters for athlete id = 1 as

$$y_1 = (765,\ 789,\ 836,\ 890)^\top$$

$$x_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

and since the binding condition $b_1 = b_2 = b_3 = b_4$ we can write

$$\boldsymbol{b}_1 = (b_0,\ b,\ b,\ b,\ b)^\top\ .$$

In the case of missing values, like for the long jump of athlete id = 60 from Table 2, we get instead

$$y_{60} = (903, 1079)^\top$$

$$x_{60} = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

and with different $b_2 = b_4$ since it is a different event

$$\boldsymbol{b}_{60} = (b_0,\ b,\ b)^\top\ .$$

3.2 Linear mixed effect models for longitudinal data

Longitudinal data analyses require serial observations on the same unit under often variating circumstances (Laird & Ware, 1982, 1). This may result in a highly unbalanced data i.e. missing values or dropouts which is common in most studies.

According to Laird and Ware (1982) it is challenging to find a general covariance structure for unbalanced data which calls for the use of random effect within the linear modelling. These kinds of models are called mixed models. A mixed model extends the use of fixed and random variables from the general mean μ (fixed) and error terms ε (random) used in most models (Gruber & Searle, 2017, 497).

The basic linear modeling works under the assumption that all observations are drawn from the same, independent and identically distributed population. In the mixed model the data has a more hierarchical structure containing levels or clusters. The observations within these levels or clusters are assumed to be dependent since they belong to the same subpopulation while the observations between levels or clusters are assumed to be independent (Demidenko, 2013, 1). One type of mixed model is the linear mixed-effect model developed by Laird and Ware (1982) that has the advantage of dealing with correlated errors within levels and hence does not require a balanced data:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{\beta}_i + \varepsilon_i \qquad (i = 1, 2, \ldots, p)$$

where the error terms $\varepsilon_i \sim N(0, R_i)$ are independent with the positive-definite covariance matrix R_i. One key feature is the unknown vector $\boldsymbol{\alpha} \in \mathbb{R}^{p\times 1}$ (intercept) representing the fixed effects of the population and the unknown vector $\boldsymbol{\beta}_i \in \mathbb{R}^{k\times 1}$ (slope) representing the specific random effects for each subject. The number of individuals (observations) are m. The unknown intercept $\boldsymbol{\alpha}$ is linked to the response vector $\boldsymbol{y}_i$ by the known design matrix $\boldsymbol{X}_i \in \mathbb{R}^{n_i\times p}$. The unknown slope $\boldsymbol{\beta}_i$ is linked to $\boldsymbol{y}_i$ by the known design matrix $\boldsymbol{Z}_i \in \mathbb{R}^{n_i\times k}$.

In our case $p = 61$ and $k = 4$ which means that $\boldsymbol{\alpha} \in \mathbb{R}^{61\times 1}$ represents the estimated score for each athlete at the first level and $\boldsymbol{\beta}_i \in \mathbb{R}^{4\times 1}$ the improvement in points on each level for each decathlete i. The variable n_i means each level where the athlete with id i has a score. In most of our cases $n_i = 4$ but as we can see from Table 2 id $= 60$ has $n_i = 2$.

Laird and Ware (1982) introduce two stages of the linear mixed-effects model (LME), one where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_i$ are considered fixed and one where $\boldsymbol{\beta}_i$ is considered random.

We obtain

$$E[\boldsymbol{y}_i] = \boldsymbol{X}_i\boldsymbol{\alpha}$$

and

$$\mathrm{Cov}[\boldsymbol{y}_i] = R_i + \boldsymbol{Z}_i D_i \boldsymbol{Z}_i^{\top} = \textstyle\sum_{\boldsymbol{i}}$$

which results in the multivariate normal distribution for the response vector:

$$\boldsymbol{y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\alpha}, \textstyle\sum_i\ ).$$

3.3 Estimation

One advantage of the linear mixed effect model is that it considers dependency within the levels or clusters and hence handles correlated errors. According to Garson (2013, 6) this gives the benefit of using maximum likelihood (ML) estimation and restricted maximum likelihood (REML) estimation which does not require the data to be balanced.

The disadvantage of the ML estimation is well presented by Efron and Morris (1975) when demonstrating the Stein´s estimator which like REML is based on empirical Bayes methods. Laird and Ware (1982) also provide good reasons for preferring REML over ML due to the common lack of information regarding the variance in the fixed effect α.

REML uses a linear transformation $\boldsymbol{LY}$ satisfying $\boldsymbol{LX} = 0$ that does not depend on α (Agresti, 2015, 306) and is the reason the model is called “restricted”. Linear transformation is common in scaling which is useful in Bayesian modeling (Gelman & others, 2014, 368).

We will assume the reader is familiar with Bayesian inference (Gelman & others, 2014, 6-8) and the ML estimation procedure and focus only on the REML estimation introduced by Patterson & Thompson in 1971.

The proof of the following methodology is detailed in Appendix A.

The maximized log-likelihood function of parameters $(\boldsymbol{\alpha}, \boldsymbol{\theta})$ has the form:

$$l(\boldsymbol{\alpha}, \boldsymbol{\theta}) = -\frac{1}{2} det \sum(\boldsymbol{\theta}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\alpha})^{\top} \sum(\boldsymbol{\theta})(\boldsymbol{y} - \boldsymbol{X\alpha})$$

where

$$\sum(\boldsymbol{\theta}) = \begin{bmatrix} \sum_1(\boldsymbol{\theta}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sum_m(\boldsymbol{\theta}) \end{bmatrix}$$

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_m \end{pmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{X}_m \end{bmatrix}$$

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix}.$$

The estimation for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and the variance parameter $\boldsymbol{\theta}$ can be derived using Bayesian interpretation of REML (Appendix A).

According to Laird and Ware (1982) and based on Harville (1976) we obtain our wanted REML-estimate of $\boldsymbol{\theta}$ noted as $\widehat{\boldsymbol{\theta}}_R$ by maximizing the limiting marginal likelihood of:

$$\boldsymbol{y}_i | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \Gamma \sim \boldsymbol{N}(A_i\boldsymbol{\alpha} + B_i\boldsymbol{\beta}_i, R_i)$$

where $A_i \in \mathbb{R}^{n_i \times p}$ and $B_i \in \mathbb{R}^{n_i \times k}$ are assumed known and

$$\boldsymbol{\alpha} \sim N(0, \Gamma)$$

$$\boldsymbol{\beta}_i \sim N(0, \Delta)$$

under the assumption, that $\Gamma^{-1} \to 0$. The matrix $R_i \in \mathbb{R}^{n_i \times n_i}$ is the unknown covariance matrix for the random noise ε_i.

In line with Laird and Ware (1982) and Chapter 6 (Theorem 1)

$$\widehat{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}_R) = \mathbb{E}(\boldsymbol{\alpha} | \boldsymbol{y}, \widehat{\boldsymbol{\theta}}_R, \Gamma^{-1}) = \Sigma_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}_R) \boldsymbol{X}_i^\top {\Sigma_{\boldsymbol{y}|\boldsymbol{\alpha}}}^{-1} \boldsymbol{y}_i.$$

Additionally (Appendix A, Theorem 2) we have

$$\widehat{\boldsymbol{\beta}_i}(\widehat{\boldsymbol{\theta}}_R) = \mathbb{E}(\boldsymbol{\beta}_i | \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}_R, \Gamma^{-1}) = \boldsymbol{D}_i \boldsymbol{Z}_i^\top \left(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}_R)\right).$$

3.4 AIC and BIC criteria

The Akaike Information Criterion (AIC) (Akaike, 1973) and the Bayesian Information Criterion (BIC) (Schwarz, 1978) are two popular methods to measure the performance of a model and to compare variables. They are like other similar methods based on a given point estimate of the fitted data and the log predictive density and work on different model designs (Gelman & others, 2014, 169).

They are defined by:

$$AIC = -2l(\hat{\alpha}|\boldsymbol{Y}) + 2K$$
$$BIC = -2l(\hat{\alpha}|\boldsymbol{Y}) + K\ln(N),$$

where K is the number of parameters we add to our model, N the number of observations and $l(\hat{\alpha}_{ML}|\boldsymbol{Y})$ the maximized log-likelihood for the model. As we add parameters to our model we want it to be as simple as possible i.e. to contain as few parameters as possible. The maximized log-likelihood function decreases when the model improves. The BIC is a little restrained gaining a higher value than AIC due to its penalty term application $\ln(N)$ that get larger with more observations. Compared to another model lower AIC and BIC indicates the better model but the value itself does not have an interpretation (Busemeyer & Diederich, 2014).

3.5 Example

As a demonstrating example the linear model and the linear mixed effect model have been compared to estimate the development in pole vault over the years based on the five last world record holders in the decathlon. We note that here we have $n_i = 1, \dots, 13$ which is a larger number than eventually in our official analysis where $1 < n_i \leq 4$. The benefit of using levels instead of years is visible since someone has peaked at year 6 while someone else has done it at year 13. The `lm` and `lme` functions from R´s `nlme` -package have been used and the models likelihood of fit have been compared using the `anova.lme` (Analysis of Variance) function.

Table 4. R-script.

```
model1 <- lm(PVP ~ year, data = wr, na.action = "na.omit")

model2 <- lme(PVP ~ year, data = wr, na.action = "na.omit", random = ~1|id)


> coef(model1)
(Intercept)        year
  609.26422    32.79305

> coef(model2)
   (Intercept)  year
7     725.4102 42.1595
16    609.5682 42.1595
30    646.4295 42.1595
45    501.6923 42.1595
51    395.4854 42.1595

> anova.lme(model1,model2)
       Model df      AIC      BIC    logLik   Test L.Ratio
model1     1  3 619.3239 624.8743 -306.6619
model2     2  4 582.3015 589.7021 -287.1508 1 vs 2 39.0224
       p-value
model1
model2  <.0001
```

As we can see from Table 4 the ANOVA-test provides lower AIC and BIC for the linear mixed-effect model showing it´s a better fit and significantly different from the linear model due to the low p-value. The linear model provides only one intercept and one slope while the mixed model predicts five different intercepts for each individual but also one slope because no random effect was put into the model.

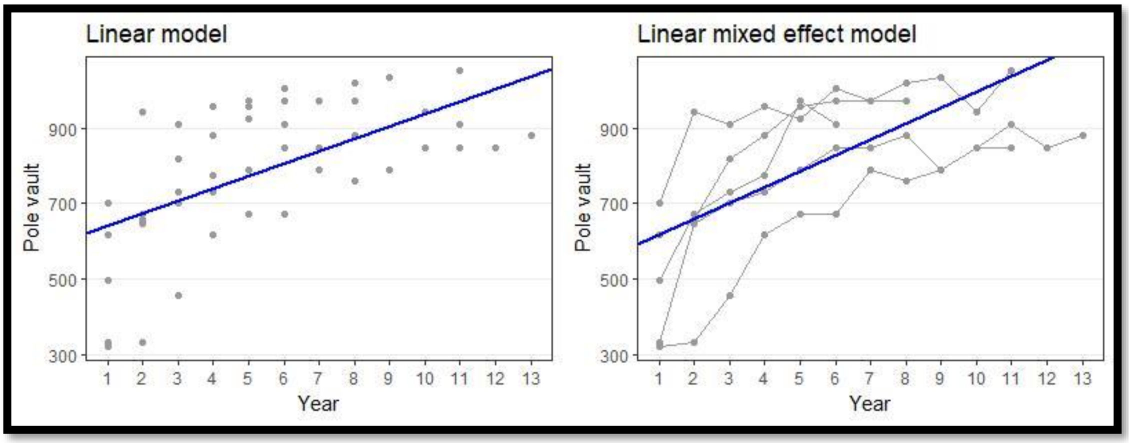


Figure 2. Linear model versus linear mixed effect model.

In Figure 2 both models are drawn to visualize the difference. In the linear model you can see the effect from the unequal number of years affecting the slope. In the linear mixed effect model the individual development is observed resulting in a steeper slope.

4 RESULTS

In this section, we study the correlations between events for decathletes under 8 000 points and over 8 000 points. In his unpublished Bachelor´s thesis (Ylöstalo, 2018) the author started using Pearson´s correlation coefficient *r* for comparing the correlation between events for decathletes under 8 000 points and over 8 000 points. The following results are obtained using the same method and the *r* will refer to the Pearson's correlation coefficient. Figure 3 shows the correlation between the results in groups 1 and 2 (under 8000p) and groups 3 and 4 (over 8000p) the correlation between the shot put and the discus throw is strong in both cases (r = 0.69 for under 8000p and r = 0.65 for over 8000p). In groups 1 and 2 the long jump holds the strongest correlation to the total score (r = 0.59) while the corresponding correlation for groups 3 and 4 comes from the 110m hurdles (r = 0.49).

In groups 1 and 2 the high jump and the 1500m correlates to the total score (r = 0.30 and r = 0.23) while this changes in groups 3 and 4 (r = 0.01 and r = -0.01). There are more negative correlations between events in groups 3 and 4.

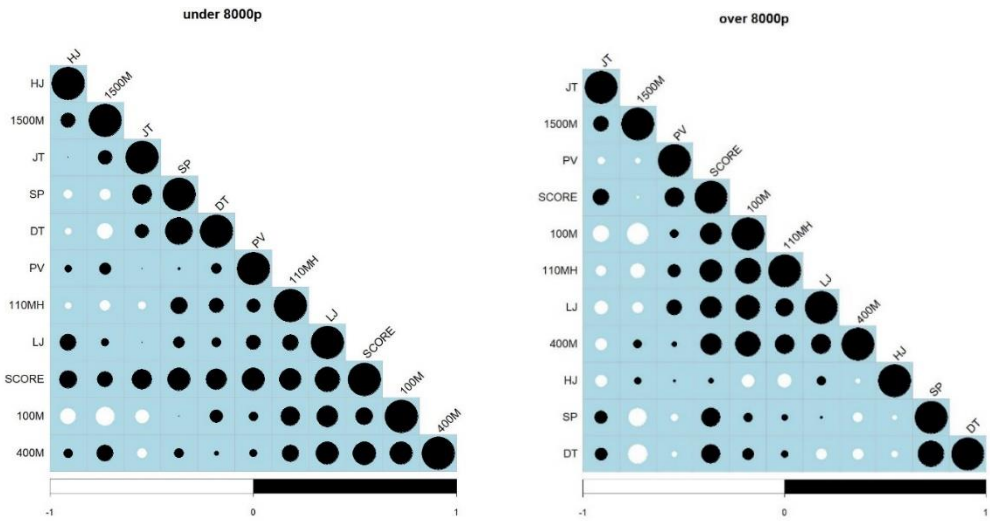


Figure 3. Correlations for the decathletes under 8000 points and over 8000 points.

4.1 Development in each event

When comparing the percentage development in points (due to differences in the total amount of points) through linear mixed effect modeling for each level (Figure 4). All the events show improvement to level 3. From level 3 to level 4 all events except for the high jump and the 1500m improve. The high jump stays equal while the 1500m decreases slightly. The pole vault shows the biggest improvement overall with a 26.6% improvement to the third level and a total progress of 32.8%. The sprint-based events (100m, 400m and 100m hurdles) show a similar pattern to each other.

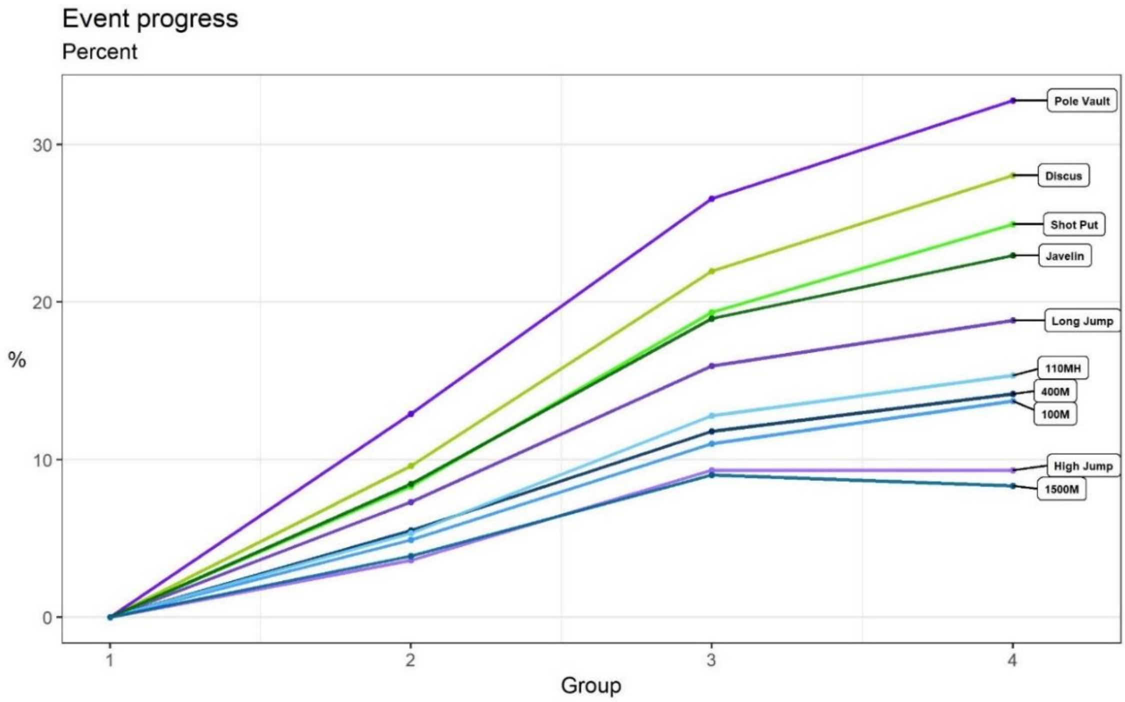


Figure 4. The overall percentage development in each event per level (group) based on the starting value.

4.2 Development based on initial level

In Figure 5 four events are presented individually based on the linear mixed effect modelling for three subgroups. The subgroups are picked based on the quartiles at the initial level: beyond average (75%+), average (25%-75%) and below average (25%-). From each group the median of each stage is presented (Figure 5).

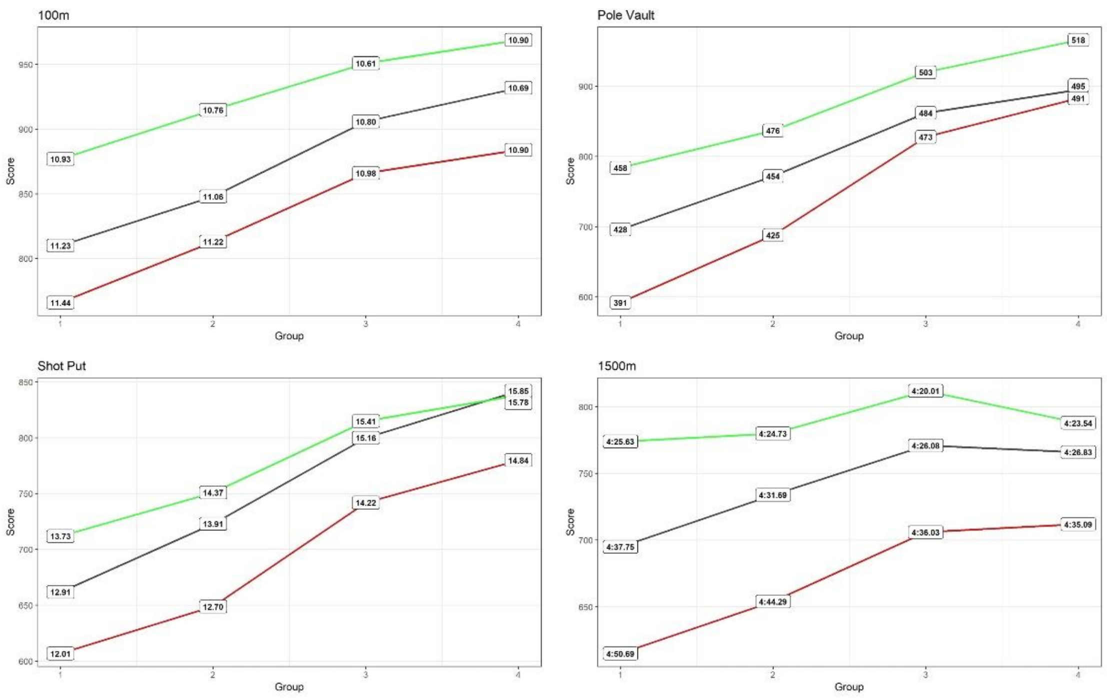


Figure 5. The median development of three subgroups based on their starting level. Strong performers in level 1 are green and corresponding weak are red and the rest are black. The result matching the points are labeled.

The 100m shows a similar development for each group while the weaker subgroup is developing more in the pole vault. The average subgroup shows the best development in the shot put and surpasses the stronger group in the final stage. In the 1500m there are big differences in the starting point. The strong group doesn´t improve after the third level while the other subgroups stay approximately within one second. In all three events except for the pole vault the weaker subgroup also stays the weakest during all stages.

5 DISCUSSION

5.1 Sprints are key

One of the aims of this study was to find the most important feature for a decathlete. Figure 1 shows, that the sprints and the jumps are giving more points than the throws so the natural staring point would be to start from the 100m. The 100m correlates strongly to the long jump, 400m and 110m hurdles over the whole period of study but at a later stage during the athlete's best years this correlation is bigger.

In the linear mixed effect model this can be interpreted from the similar development the sprint-based events hold to each other. Even though the sprinting events are continuously developing their impact on the total score becomes smaller. One explanation to this could be, that sprints are a big part of the everyday training and not as technically demanding as the jumps and throws. To become an elite decathlete all the events need to be on a high level, but this could very well indicate that the difference between this elite and the rest lies in the possibility to develop your throws and jumps beside this clear importance of the sprints.

5.2 The long jump and the 400m

The long jump and the 400m are the two events where the strong group develop more at the final stage which differs from the normal pattern for single event athletes in the same events (Fung & Ha, 1994).

One reason for this happening in the long jump could be explained by the reasonably common fact that many decathletes jump close to or beyond their personal best in this event during their best series because of the relatively high amount of points coming from each centimeter in this event compared to many other.

This explanation would not hold for the 400m since the relatively amount of points coming from each second in this event is rather small. The reason for this might be more spiritual because motivation is an important factor in the decathlon (Edourad & others, 2010). A decathlete in top shape gets this confirmed during the whole first day where most of the events based on pure physical capacity are measured and the 400m wraps up that day. Noticeable is also the fact that the initially weak runners gain a lot of progress during the early stages of their career and that decathletes during the final stage are relatively close to each other.

5.3 Strong high jumpers peak earlier

When comparing the development in the high jump it is noticeable, even though it is not by a lot, that the initially strong high jumpers peak before they break 8 500 points. The development after 8 000 points is not big in any group, but especially the initially weak jumpers gain a lot during the earliest stage.

5.4 The second day

The second day is known for its more technical demand, which gives more room for error regardless of physical shape. The second day is known for its dramatical turns on the competition where strong competitors might fail and lose points.

In his unpublished Bachelor´s thesis (Ylöstalo, 2018) the author noticed that the pole vault did not have a clear correlation to the other events since the Pearson correlation coefficient *r* ranged from 0 to 0.15. One reason could be the challenge of being the 8th event performed with a tired body as well as being perhaps the most technically demanding event of the decathlon. The pole vault also stands out as the most advancing event on the journey over 8 500 points. When studying the different patterns from the three different levels of starting point this could be explained by the fact, that the weakest pole vaulters from the beginning make a huge progress. According to this study a good and a bad pole vaulter is in general only separated by only 27 centimeters in the end compared to almost 70 centimeters in the beginning.

5.5 Difference in correlation

The strong negative correlation between the 1 500m and the discus throw and the shot put increases during these athlete´s best years which also supports the results in the author´s unpublished Bachelor´s thesis (Ylöstalo, 2018). However, in the earlier study the javelin throw seemed to hold a negative correlation to the 1 500m with a Pearson correlation coefficient *r* of -0.09 for decathletes below 8 000 points and *r* of -0.11 for decathletes above 8 000 points. The result in this study is different with the corresponding Pearson correlation coefficients *r* of 0.20 and 0.23. This observation could be explained by the fact that the javelin throw is more based on running and a different kind of footwork and hence supports the theory that a strong running base is of importance to maintain a good overall physique during the final events of the decathlon.

5.6 Weak runners are developing

The 1500m may be the most mentally demanding event in the decathlon since it sums the whole decathlon and all athletes know before going into the race how fast they need to run to make a certain score or if they are only running for a position, which is common especially in major championships. When taking physical condition in consideration the 1500m may play an even bigger role since the athlete need a certain standard of aerobic ability to recover not only between events during the decathlon but, also during training throughout the year (Bompa, 1999)(Martin & Coe, 1997, 153). If this is conscious remains unclear, but the decathletes with a weaker capacity in the 1 500m from the beginning develop more than the other decathletes who are average or beyond at an early stage. This development is similar in the 400m and it can be assumed that it contains partially the same individuals due to the relation between speed endurance and middle-distance running (Duffield etc., 2005).

There are probably multiple components affecting this phenomenon, but a few very likely affecting factors might be better training or just the fact that it is easier to develop a weakness than a strength. Another feature of the 1 500m is the tactics and pace of running the distance which require different mental skills compared to power performances (Butovskaya & others, 2017). An athlete doesn´t necessarily know how to run at his full capacity at a younger age and develop this skill later in the career.

5.7 Conclusion

Since this study has only compared decathlete´s best series per season there might be occasions when some events go better or worse than considered normal for the athlete. A future study could take all the decathlons as well as single event starts into consideration to give an even more complex picture of the decathlete´s profile. After getting introduced to linear mixed effect modeling and all the possibilities it could also be possible to instead of levels analyze years. Then a good model would estimate the initial level for the athletes lacking early marks. Another even more rewarding thing would be to by mixing linear models and then use for instance factor analysis on how progress in certain events might affect other events negatively.

Since the breakthrough of the computer science and the internet there are better known facts and competitions from today´s athletes than back in the 70´s and 80´s where many results to this study are taken. The aim of this study was to perform a statistical analysis on this topic which could be used as inspiration for coaches, athletes and federations to gain a deeper understanding to one of the most multifaceted events in the world.

There are many things leading up to a good decathlon and there will always be exceptions to the rule, which is something every coach and athlete must take into consideration before making any final conclusions. After observing the progress among the initially weak runners on 400 meters and 1 500 meters it could wrap up this study in recommending a good and wide running base for all who want to become a great decathlete.

REFERENCES

Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Hoboken, New Jersey: Wiley.

Akaike, H., (1973). *Information theory and an extension of the maximum likelihood principle*. Acadeemiai Kiadi, Budapest, Hungary. International Symposium on Information Theory, 2nd ed.: 267–281.

Bauersfeld, K.H. & Schröter, G. (1989). *Grundlagen der Leichtathletik*. Berlin, Germany. Suom. Oikarinen, L. Yleisurheiluvalmennuksen perusteet.

Bompa, T.O. (1999). *Periodization: Theory and Methodology of Training 4th edition*. Human Kinetics, USA.

Butovskaya, M., Chamari, K., Messaoud, T., Miarka, B., Siala, H., Slimani, M., Souissi, N. & Znazen, H. (2017). *Mental skills comparison between elite sprint and endurance track and field runners according to their genetic polymorphism: a pilot study*. The Journal of sports medicine & physical fitness Vol 57(9):1217-1226.

Busemeyer, J. R. & Diederich A. (2014). *Estimation and Testing of Computational Psychological Models.* Neuroeconomics, Second Edition: 49-61.

Demidenko, E. (2013). *Mixed models: theory and applications with R, 2nd edition*. Hoboken: Wiley 2013.

Duffield, R., Dawson, B. & Goodman, C. (2005). *Energy system contribution to 400-metre and 800-metre track running*. Journal of Sports Science Vol 23(3): 299-307.

Edourad, J. L., Edourad, P., Morin, J. B. & Pruvost, J. (2010). *Causes of dropouts in decathlon. A pilot study*. Phys Ther Sport 11(4): 133-135.

Efron, B. & Morris, C. (1975). *Data analysis using Stein's estimator and its generalizations*. Journal of the American Statistical Association Vol 70: 311-319.

Etcheverry, S.G. (1995). *Overview 2: Profile of the Decathlete*. New Studies in Athletics: 23-27.

Fung, L. & Ha, A. (1994). *Changes in track and field performance with chronological aging*. International Journal of aging & human development Vol 38(2): 171-180.

Garson, G. D. (2013). *Hierarchical Linear Modeling: Guide and Applications*. Los Angeles: SAGE.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2014). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/Crc Texts in Statistical Science. Boca Raton, Fla.: CRC 2014.

Harville, D. A. (1976). *Extension of the Gauss-Markov theorem to include the estimation of random effects.* Annals of Statistics Vol(4): 384-395.

Harville, D. A. (1977). *Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems*. Journal of the American Statistical Association Vol 72(358): 320-338.

Karvonen, M.J. & Niemi, M. (1953). *Factor Analysis of Performance in Track and Field Events*. Arbeitsphysiologie. 127-133.

Laird, N. M. & Ware, J. H. (1982). *Random effects models for longitudinal data: an overview of recent results*. Biometrics Vol 38(4): 963-974.

Martin, D. E. & Coe, P. N. (1997). *Better training for distance runners*. Human Kinetics.

Patterson, H. D. & Thompson, R. (1971). *Recovery of inter-block information when block sizes are unequal*. Biometrika Vol 58(3):545–554

Schwarz, G. (1978). *Estimating the dimension of a model*. Annals of Statistics Vol 6(2): 461-464.

Searle, S. R. & Gruber, M. H. J. (2017) *Linear Models, Second Edition*. Hoboken, New Jersey: Wiley 2017.

Stellingwerff, T., Maughan, R.J. & Burke, L.M. (2011). *Nutrition for Power Sports: Middle-distance Running, Track Cycling, Rowing, Canoeing/Kayaking and Swimming*. Journal of Sports Sciences. 29 (1): 79-89.

Wang, Z. & Lu, G. (2007). *The Czech Phenomenon of Men´s Decathlon Development*. International Journal of Sports Science and Engineering. 209-214.

Westera, W. (2006). *Decathlon: Towards a balanced and sustainable performance assessment method*. New Studies in Athletics. 39-51.

Ylöstalo, O. (2018). *Twenty years of decathlon*. University of Helsinki: Bachelor's thesis. Unpublished.

Zarnowski, F. (1989). *The Decathlon*. Leisure Press, USA.

Decathlon 2000 (2019). www.decathlon2000.com/athletes, cited 12.9.2019

IAAF Council, (2001). IAAF Scoring Tables for Combined Events. 10 (2), 23-27

Appendix A

We recall that the response vector in a linear mixed effect (LME) model follows a full multivariate normal distribution

$$\boldsymbol{y}_i \sim N(\boldsymbol{\mu}_i, \textstyle\sum_i) \qquad (1)$$

where $i = 1, \dots, m$ and $N = \sum_{i=1}^{m} n_i$.

$$\boldsymbol{\mu}_i = \boldsymbol{X}_i \boldsymbol{\alpha} \qquad (2)$$

$$\textstyle\sum_i = \begin{bmatrix} \sum_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sum_m \end{bmatrix} \qquad (3)$$

where $\boldsymbol{\mu}_i$ is a $n_i \times 1$ vector and $\sum_i$ a $n_i \times n_i$ matrix. This leads to

$$\boldsymbol{y} \sim \boldsymbol{N}_{\boldsymbol{N}}(\boldsymbol{\mu}, \textstyle\sum)$$

where $\boldsymbol{\mu}$ is a $N \times 1$ vector and $\sum_i$ a $N \times N$ matrix.

Initially we also present the ideas introduced by Harville (1977) where you have two different stages of LME models. Let’s call them stage 1 and stage 2. We recall the LME formula:

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\alpha} + \boldsymbol{Z}_i \boldsymbol{\beta}_i + \varepsilon_i$$

Stage 1. In stage 1 both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_i$ are considered fixed while $\varepsilon_i \sim N(0, R_i)$ is independent with the positive-definite covariance matrix R_i.

Stage 2. In stage 2 $\boldsymbol{\alpha}$ and ε_i are similar with stage 1 but $\boldsymbol{\beta}_i$ is considered random and independent of each other and $\boldsymbol{\varepsilon}_{\boldsymbol{i}}$. The distribution of $\boldsymbol{\beta}_{\boldsymbol{i}}$ is $N(0, D_i)$ where $D_i \in \mathbb{R}^{k \times k}$ is positive-definite.

I.e. the stage 1 model works under the assumption of homogeneity between individuals and hence fits into a hierarchical model. Stage 2 works under the assumption of heterogeneity between individuals and a random unknown model.

Laird and Ware (1982) explain the variance parameter $\boldsymbol{\theta}$ as a q-vector of variance and covariance parameters found in R_i and D_i for $i = 1, \dots, m$ and $D_i \in \mathbb{R}^{k \times k}$.

Example

If $n_i = 1$ for every i then $N = m$ which is followed by that R_i is the variance of ϵ_i marked as θ_i. If $k = 2$ then due to symmetry

$$D = \begin{bmatrix} \theta_{m+1} & \theta_{m+2} \\ \theta_{m+2} & \theta_{m+3} \end{bmatrix}$$

which is followed by that $\boldsymbol{\theta} = (R_1, \dots, R_m, D) = (\theta_1, \dots, \theta_m, \theta_{m+1}, \theta_{m+2}, \theta_{m+3})$.

If k for D_i is still 2 and $n_1 = 2, n_2 = n_3 = \cdots = n_m = 1$ then $N = m + 1$ and we obtain $R_1 \in \mathbb{R}^{2\times 2}$.

Hence $\boldsymbol{\theta} = (R_1, R_2, R_3 \dots, R_m, D) = (\theta_{11}, \theta_{12}, \theta_{13}, \theta_2, \theta_3 \dots, \theta_m, D_{D_1}, D_{D_2}, D_{D_3})$.

Lemma 1

Under the assumptions (1)-(3) the likelihood function of parameters $(\boldsymbol{\alpha}, \boldsymbol{\theta})$ is

$$L(\boldsymbol{\alpha}, \boldsymbol{\theta}) = \prod_{i=1}^{m} \det\left(\Sigma_i(\boldsymbol{\theta})\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left((\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha})^\top \Sigma_i(\boldsymbol{\theta})^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha})\right)\right)$$

and especially the log-likelihood function is

$$\begin{aligned} l(\boldsymbol{\alpha}, \boldsymbol{\theta}) &= -\frac{1}{2}\det\Sigma(\boldsymbol{\theta}) - \frac{1}{2}\sum_{i=1}^{m}\left((\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha})^\top \Sigma_i(\boldsymbol{\theta})^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha})\right) \\ &= -\frac{1}{2}\det\Sigma(\boldsymbol{\theta}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha})^\top \Sigma(\boldsymbol{\theta})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}) \end{aligned}$$

where

$$\Sigma(\boldsymbol{\theta}) = \begin{bmatrix} \Sigma_1(\boldsymbol{\theta}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \Sigma_m(\boldsymbol{\theta}) \end{bmatrix}$$

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_m \end{pmatrix}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{X}_m \end{bmatrix}$$

and

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_m \end{pmatrix}$$

Computation and estimation through Bayesian inference

In this section we introduce how the estimation for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be derived using Bayesian interpretation of REML. Initially we rewrite the previously processed two stages in a Bayesian formulation (Laird and Ware, 1982, 6-7).

At stage 1 $\boldsymbol{y}_i$ depends only on $\boldsymbol{\alpha}$ and is conditionally independent of $\boldsymbol{\theta}$ given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_i$, i.e. $\boldsymbol{y}_i \perp\!\!\!\perp \{\boldsymbol{\beta}_j; j \neq i\}$ and $\boldsymbol{y}_i \perp\!\!\!\perp \boldsymbol{\theta}|(\boldsymbol{\alpha}, \boldsymbol{\beta}_i)$. This gives the conditional probability

$$\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta} \sim \boldsymbol{N}(\boldsymbol{X}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{\beta}_i, R_i).$$

At stage 2, where $\boldsymbol{\alpha} \sim N(0, \Gamma)$ and $\boldsymbol{\beta}_i \sim N(0, D)$ are independent and $\text{cov}(\boldsymbol{\alpha}, \boldsymbol{\beta}_i) = 0$ for $i = 1, \dots, m$, we get the probability

$$\boldsymbol{y}_i \sim \boldsymbol{N}(0, \boldsymbol{X}_i\boldsymbol{\Gamma}\boldsymbol{X}_i^\top + \boldsymbol{Z}_i\mathbf{D}\boldsymbol{Z}_i^\top + R_i).$$

In line with Laird and Ware (1982) we let $\boldsymbol{\theta}$ represent the unknown parameters in R_i and D while Γ represents the variation within the population.

From fully Bayesian model to REML

In our model we have the observations $\boldsymbol{y}_i \in \mathbb{R}^{n_i}$ $(N = \sum n_i,\ i = 1, \dots, m)$ and the unknown parameters $\boldsymbol{\alpha} \in \mathbb{R}^p$, $\boldsymbol{\beta}_i \in \mathbb{R}^k$, $\boldsymbol{\theta}$, Γ and the “random noise” $\varepsilon_i \in \mathbb{R}^{n_i}$.

The model:

$$\boldsymbol{y}_i = A_i\boldsymbol{\alpha} + B_i\boldsymbol{\beta}_i + \varepsilon_i$$

where both $A_i \in \mathbb{R}^{n_i \times p}$ and $B_i \in \mathbb{R}^{n_i \times k}$ are assumed known.

The random noise $\varepsilon_i|(\boldsymbol{\theta}, \Gamma)$ is assumed to be $N(0, R_i)$ distributed and independent for every i where R_i is an unknown $n_i \times n_i$ covariance matrix. Following the conditional independence $(\varepsilon_i, \boldsymbol{\alpha}, \boldsymbol{\beta}_i) \perp\!\!\!\perp |(\boldsymbol{\theta}, \Gamma)$ we obtain

$$\boldsymbol{y}_i|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta},\Gamma \sim \boldsymbol{N}(A_i\boldsymbol{\alpha} + B_i\boldsymbol{\beta}_i, R_i).$$

According to Laird and Ware (1982) we can obtain an estimate of $\boldsymbol{\theta}$ based on the information we have in the data but not an estimate of Γ. As Harville (1976) indicates, we can obtain an estimate of $\boldsymbol{\theta}$ by maximizing the limiting marginal likelihood equivalent to the REML-estimate by letting $\Gamma^{-1} = 0$.

When $\boldsymbol{\alpha}{\sim}N(0,\Gamma)$ and $\boldsymbol{\beta}_i{\sim}N(0,\Delta)$ are independent with the unknown covariance matrices $\Gamma \in R^{p\times p}$and $\Delta \in R^{k\times k}$ the conditional distributions $\boldsymbol{\alpha}|(\boldsymbol{\theta},\Gamma){\sim}N(0,\Gamma)$ and $\boldsymbol{\beta}_i|(\boldsymbol{\theta},\Gamma){\sim}N(0,\Delta)$ follow. Hence, as $\Gamma^{-1} \to 0$, we can define $\boldsymbol{\theta}$ as an unknown *q*-dimensional vector (Δ, ε_i) which is our wanted REML-estimate.

To be able to calculate the posterior means of the conditional $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_i$ we first need to prove the following.

Lemma 2

If $W \in R^n$ and $T \in R^m$ are random vectors satisfying

$$W|T \sim N(\mu_W + AT, \Sigma_W) \text{ and } T \sim N(\mu_T, \Sigma_T)$$

with the constant vectors $\mu_W \in R^n$ and $T \in R^m$ and a constant matrix $A \in R^{m\times n}$ the conditional distribution of T given W is

$$T|W \sim N(\mu_{T|W}, \Sigma_{T|W}),$$

where

$$\Sigma_{T|W} = \left(A^\top \Sigma_W^{-1} A + \Sigma_T^{-1}\right)^{-1} \text{ and} \tag{4}$$

$$\mu_{T|W} = \Sigma_{T|W}\left(A^\top \Sigma_W^{-1}(w - \mu_W) + \mu_T \times \Sigma_T^{-1}\right)^{-1}. \tag{5}$$

Proof

We will show the claim using Bayes Theorem.

Since for fixed $w \in R^n$ the conditional density of W given T is proportional

$$p_{W|T}(w|T) \propto \exp\left(-\tfrac{1}{2}\big(w-(\mu_W+AT)\big)^\top \Sigma_W^{-1}(w-(\mu_W+AT))\right)$$

and the prior density of T is proportional to

$$p_T(T) \propto \exp\left(-\tfrac{1}{2}(T-\mu_T)^\top \Sigma_T^{-1}(T-\mu_T)\right)$$

we obtain with Bayes Theorem, that the conditional density for T given $W = w$ is for fixed $w \in R^m$ proportional to

$$p_{T|W}(T|w) \propto \exp\left(-\frac{1}{2}\big(T^\top A^\top \Sigma_W^{-1} AT + T^\top \Sigma_T^{-1} T - 2T^\top A^\top \Sigma_W^{-1}(w-\mu_W) - 2T^\top \Sigma_T^{-1}\mu_T\big)\right.$$

$$= \exp\left(-\tfrac{1}{2}\big(T^\top\big(A^\top \Sigma_W^{-1} A + \Sigma_T^{-1}\big)T - 2T^\top\big(A^\top \Sigma_W^{-1}(w-\mu_W) + \Sigma_T^{-1}\mu_T\big)\big)\right).$$

With the values obtained from (4) and (5) we can rewrite the previous conditional density as proportional to

$$p_{T|W}(T|w) \propto \exp\left(-\frac{1}{2}\big(T^\top \Sigma_{T|W}^{-1} T - 2T^\top \Sigma_{T|W}^{-1}\mu_{T|W}\big)\right)$$

$$\propto \exp\left(-\tfrac{1}{2}\big(T-\mu_{T|W}\big)^\top \Sigma_{T|W}^{-1}\big(T-\mu_{T|W}\big)\right). \qquad \square$$

From this we can derive the distribution of $(\boldsymbol{\alpha},\boldsymbol{\beta})|\boldsymbol{y}_i,\boldsymbol{\theta},\Gamma$ and compute $\mathbb{E}(\boldsymbol{\alpha}|\boldsymbol{y},\boldsymbol{\theta},\Gamma)$

and $\mathbb{E}(\boldsymbol{\beta}_i|\boldsymbol{y}_i,\boldsymbol{\theta},\Gamma)$ using matrix algebra but also in a simpler way.

Lemma 3

The condition distribution $(\boldsymbol{y}_i,\boldsymbol{\alpha},\boldsymbol{\beta}_i)|\boldsymbol{\theta},\Gamma$ is a multinormal distribution.

Proof

The conditional density of $\boldsymbol{y}_i,\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_i$ given $\boldsymbol{\theta}$ and Γ is proportional to

$$f(\boldsymbol{y}_i,\boldsymbol{\alpha},\boldsymbol{\beta}_i|\boldsymbol{\theta},\Gamma) \propto f(\boldsymbol{y}_i|\boldsymbol{\alpha},\boldsymbol{\beta}_i,\boldsymbol{\theta},\Gamma) \times p(\boldsymbol{\alpha}|\boldsymbol{\theta},\Gamma)p(\boldsymbol{\beta}_i|\boldsymbol{\theta},\Gamma)$$

which, after similar derivations like in Lemma 2, is proportional to

$$\exp\left(-\frac{1}{2} V_i \widehat{\Sigma}^{-1} V_i^{\top}\right)$$

for some covariance matrix $\widehat{\Sigma}$ and where $V_i = (\boldsymbol{y}_i \quad \boldsymbol{\alpha} \quad \boldsymbol{\beta}_i)$. □

To be able to fulfill the posterior distribution we also need to show how the observations $\boldsymbol{y}_i$ are affected by the (assumed) unknown parameters.

Lemma 4

The sampling distribution of $\boldsymbol{y}_i$ given $\boldsymbol{\alpha}, \boldsymbol{\theta}$ and Γ is normal

$$\boldsymbol{y}_i | \boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma \sim N(\boldsymbol{X}_i\boldsymbol{\alpha}, \boldsymbol{R}_i + \boldsymbol{Z}_i\boldsymbol{D}_i\boldsymbol{Z}_i^{\top})$$

Proof

From Lemma 2 we know that $(\boldsymbol{y}_i, \boldsymbol{\alpha})|\boldsymbol{\theta}, \Gamma$ is normally distributed and that $\boldsymbol{\alpha}|(\boldsymbol{\theta}, \Gamma)$ is normally distributed, which makes $\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma$ normally distributed as well. To be able to derive the posterior distribution of $\boldsymbol{\alpha}$ given $\boldsymbol{y}_i$ we first need to compute the mean and covariance of $\boldsymbol{y}_i$ given $\boldsymbol{\alpha}, \boldsymbol{\theta}$ and Γ.

$$\mathbb{E}(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma)$$

$$\mathrm{Cov}(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma).$$

After straight forward derivation we obtain

$$\mathbb{E}(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma) = \mathbb{E}(\mathbb{E}(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}_i, \boldsymbol{\theta}, \Gamma)|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma) = \mathbb{E}(\boldsymbol{X}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma) = \boldsymbol{X}_i\boldsymbol{\alpha}$$

and

$$\mathrm{Cov}(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}_i, \boldsymbol{\theta}, \Gamma) = \boldsymbol{R}_i$$

$$\mathrm{Cov}(\boldsymbol{X}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma) = \mathrm{Cov}(\boldsymbol{Z}_i\boldsymbol{\beta}|\boldsymbol{\theta}, \Gamma) = \boldsymbol{Z_i}\boldsymbol{D}_i\boldsymbol{Z}_i^{\top}.$$

Which leads to

$$\mathrm{Cov}(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma) = \mathbb{E}(\boldsymbol{R}_i|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma) + \mathrm{Cov}(\mathbb{E}(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}_i, \boldsymbol{\theta}, \Gamma)|\boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma)$$

$$= \boldsymbol{R}_i + \boldsymbol{Z_i}\boldsymbol{D}_i\boldsymbol{Z}_i^{\top}$$

which proves normality of the distribution. □

From here we can prove our first Theorem, which is how to obtain some estimate for parameter $\boldsymbol{\alpha}$.

Theorem 1

The conditional posterior distribution of $\boldsymbol{\alpha}$ given $\boldsymbol{y}_i, \boldsymbol{\theta}$ and Γ is

$$\boldsymbol{\alpha}|(\boldsymbol{y}_i, \boldsymbol{\theta}, \Gamma) \sim N(\mu_{\boldsymbol{\alpha}|\boldsymbol{y}}, \Sigma_{\boldsymbol{\alpha}|\boldsymbol{y}})$$

where the covariance matrix is

$$\Sigma_{\boldsymbol{\alpha}|\boldsymbol{y}} = (\boldsymbol{X}_i^\top \Sigma_{\boldsymbol{y}|\boldsymbol{\alpha}}^{-1} \boldsymbol{X}_i + \Gamma^{-1})^{-1}$$

that depends on $\boldsymbol{\theta}$ and Γ but not $\boldsymbol{y}_i$ and the mean

$$\mu_{\boldsymbol{\alpha}|\boldsymbol{y}} = \Sigma_{\boldsymbol{\alpha}} \boldsymbol{X}_i^\top \Sigma_{\boldsymbol{y}|\boldsymbol{\alpha}}^{-1} \boldsymbol{y}_i.$$

Proof

With the information from Lemma 2 we call the following random vector

$$T = \alpha|\boldsymbol{\theta}, \Gamma \sim N(0, \Gamma)$$

and conditional distribution of W given T

$$W|T = \boldsymbol{y}_i|\alpha, \boldsymbol{\theta}, \Gamma \sim N(\boldsymbol{X}_i\boldsymbol{\alpha}, \Sigma_{\boldsymbol{y}_i|\boldsymbol{\alpha}}).$$

We obtain from Bayes Theorem that

$$\boldsymbol{\alpha}|\boldsymbol{y}_i, \boldsymbol{\theta}, \Gamma \sim N(\mu_{\boldsymbol{\alpha}}, \Sigma_{\boldsymbol{\alpha}})$$

and it gives us the wanted mean and conditional covariance:

$$\Sigma_{\boldsymbol{\alpha}|\boldsymbol{y}} = \left(A^\top \Sigma_W^{-1} A + \Sigma_\alpha^{-1}\right) = (\boldsymbol{X}_i^\top \Sigma_{\boldsymbol{y}|\alpha}^{-1} \boldsymbol{X}_i + \Gamma^{-1})^{-1}$$

$$\mu_{\boldsymbol{\alpha}} = \Sigma_{T|W}\left(A^\top \Sigma_W^{-1}(w - \mu_W) + \Sigma_T^{-1} \mu_T\right) = \Sigma_{\alpha|\boldsymbol{y}} \boldsymbol{X}_i^\top \Sigma_{\alpha|\boldsymbol{y}}^{-1} \boldsymbol{y}_i.$$ □

Corollary 1

The data contains information about $\boldsymbol{\theta}$ and it can thus be estimated, here we call $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R$. Same application doesn't work on Γ since we typically don't have information about Γ and hence cannot estimate Γ. Harville (1976) shows that an estimated $\boldsymbol{\theta}$ with a given $\boldsymbol{y}$ obtained from maximizing the limiting marginal likelihood as $\Gamma^{-1} \rightarrow 0$ is equivalent to the REML likelihood. Based on these assumptions for $\boldsymbol{\theta}$ and Γ^{-1} we can compute the empirical expected value of the conditional $\boldsymbol{\alpha}$ given $\boldsymbol{y}$, $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R$ and $\Gamma^{-1} \rightarrow 0$ as

$$\mathbb{E}\big(\boldsymbol{\alpha} \big| \boldsymbol{y}, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R, \Gamma^{-1} \rightarrow 0\big) = \Sigma_{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}_R)\boldsymbol{X}_i^{\top}\Sigma_{\boldsymbol{y}|\alpha}^{-1}\boldsymbol{y}_i$$

where the covariance is

$$\Sigma_{\boldsymbol{\alpha}}\big(\widehat{\boldsymbol{\theta}}_R\big) = \mathrm{Cov}\big(\boldsymbol{\alpha} \big| \boldsymbol{y}, \widehat{\boldsymbol{\theta}}_R, \Gamma^{-1} \rightarrow 0\big) = \left(\boldsymbol{X}_i^{\top}\Sigma_{\boldsymbol{y}|\alpha}^{-1}\boldsymbol{X}_i\right)^{-1}.$$

In a similar way we can compute the corresponding value for $\boldsymbol{\beta}_i$, however with small modifications due to the appearance of Γ in the formula. Recall that there is no assumed prior knowledge about Γ.

Lemma 5

The conditional posterior distribution of $\boldsymbol{\beta}_i$ given $\boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\theta}$ and Γ is

$$\boldsymbol{\beta}_i | \boldsymbol{y}_i, \boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma \sim N(\mu_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}}, \Sigma_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}})$$

where the conditional mean and covariance are represented as

$$\mu_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}} = \Sigma_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}}\boldsymbol{Z}_i^{\top}\boldsymbol{R}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha})$$

and

$$\Sigma_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}} = \left(\boldsymbol{Z}_i^{\top}\boldsymbol{R}_i^{-1}\boldsymbol{Z}_i + \boldsymbol{D}_i^{-1}\right)^{-1}$$

Proof

Like in Theorem 1 we rely on the information from Lemma 2 and call the following random vectors

$$T = \boldsymbol{\beta}_i | \boldsymbol{\alpha}, \boldsymbol{\theta}, \Gamma \sim N(0, \boldsymbol{D}_i)$$

and

$$W|T = \boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}_i, \boldsymbol{\theta}, \Gamma \sim N(\boldsymbol{X}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{\beta}_i, \boldsymbol{R}_i).$$

We can rely on Lemma 2 and prove that $\Sigma_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}}$ only depends on $\boldsymbol{\theta}, \Gamma$ and that $\mu_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}}$ only depends on $\boldsymbol{y}_i, \boldsymbol{\alpha}$ and $\boldsymbol{\theta}$. This indicates that the empirical posterior $\boldsymbol{\beta}_i$ can be obtained with help of $\widehat{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}_R)$ presented in Corollary 1.

Theorem 2

Based on Corollary 1 we assume $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R$ and $\Gamma^{-1} \to 0$ and get the conditional posterior distribution of $\boldsymbol{\beta}_i$ given $\boldsymbol{y}_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R$ and Γ is

$$\mathbb{E}\left(\boldsymbol{\beta}_i \middle| \boldsymbol{y}_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R\,, \Gamma^{-1} \to 0\right) = \mathbb{E}\left(\boldsymbol{\beta}_i \middle| \boldsymbol{y}_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R, \boldsymbol{\alpha} = \widehat{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}_R)\right) = \boldsymbol{D}_i\boldsymbol{Z}_i^\top(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\alpha}}(\widehat{\boldsymbol{\theta}}_R))$$

where

$$\widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_R\right) = \mathbb{E}\left(\boldsymbol{\alpha} \middle| \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}_R, \Gamma^{-1}\right).$$

Proof

Now, in line with Lemma 5 we obtain the posterior mean of $\boldsymbol{\beta}_i$ given $\boldsymbol{y}_i, \boldsymbol{\theta}$ and Γ as

$$\begin{aligned}
\mathbb{E}(\boldsymbol{\beta}_i|\boldsymbol{y}_i, \boldsymbol{\theta}, \Gamma) &= \mathbb{E}(\mathbb{E}(\boldsymbol{\beta}_i|\boldsymbol{\alpha}, \boldsymbol{y}_i, \boldsymbol{\theta}, \Gamma)|\boldsymbol{y}_i, \boldsymbol{\theta}, \Gamma) \\
&= \mathbb{E}\left(\Sigma_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}}\boldsymbol{Z}_i^\top\boldsymbol{R}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha}) \middle| \boldsymbol{y}_i, \boldsymbol{\theta}, \Gamma\right) \\
&= \Sigma_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}}\boldsymbol{Z}_i^\top\boldsymbol{R}_i^{-1}\mathbb{E}\left((\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha}) \middle| \boldsymbol{y}_i, \boldsymbol{\theta}, \Gamma\right) \\
&= \Sigma_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}}\boldsymbol{Z}_i^\top\boldsymbol{R}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\mathbb{E}(\boldsymbol{\alpha}|\boldsymbol{y}_i, \boldsymbol{\theta}, \Gamma)).
\end{aligned}$$

Hence, when $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R$ and $\Gamma^{-1} \to 0$ we get a posterior conditional mean of $\boldsymbol{\beta}_i$ given $\boldsymbol{y}_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R$ and $\widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_R\right)$ from

$$\mathbb{E}\left(\boldsymbol{\beta}_i \middle| \boldsymbol{y}_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R\,, \Gamma^{-1} \to 0\right) = \Sigma_{\boldsymbol{\beta}|\boldsymbol{\alpha},\boldsymbol{y}}\boldsymbol{Z}_i^\top\boldsymbol{R}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_R\right)\right) = \mathbb{E}\left(\boldsymbol{\beta}_i \middle| \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}_R, \widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_R\right)\right).$$

This implies, that

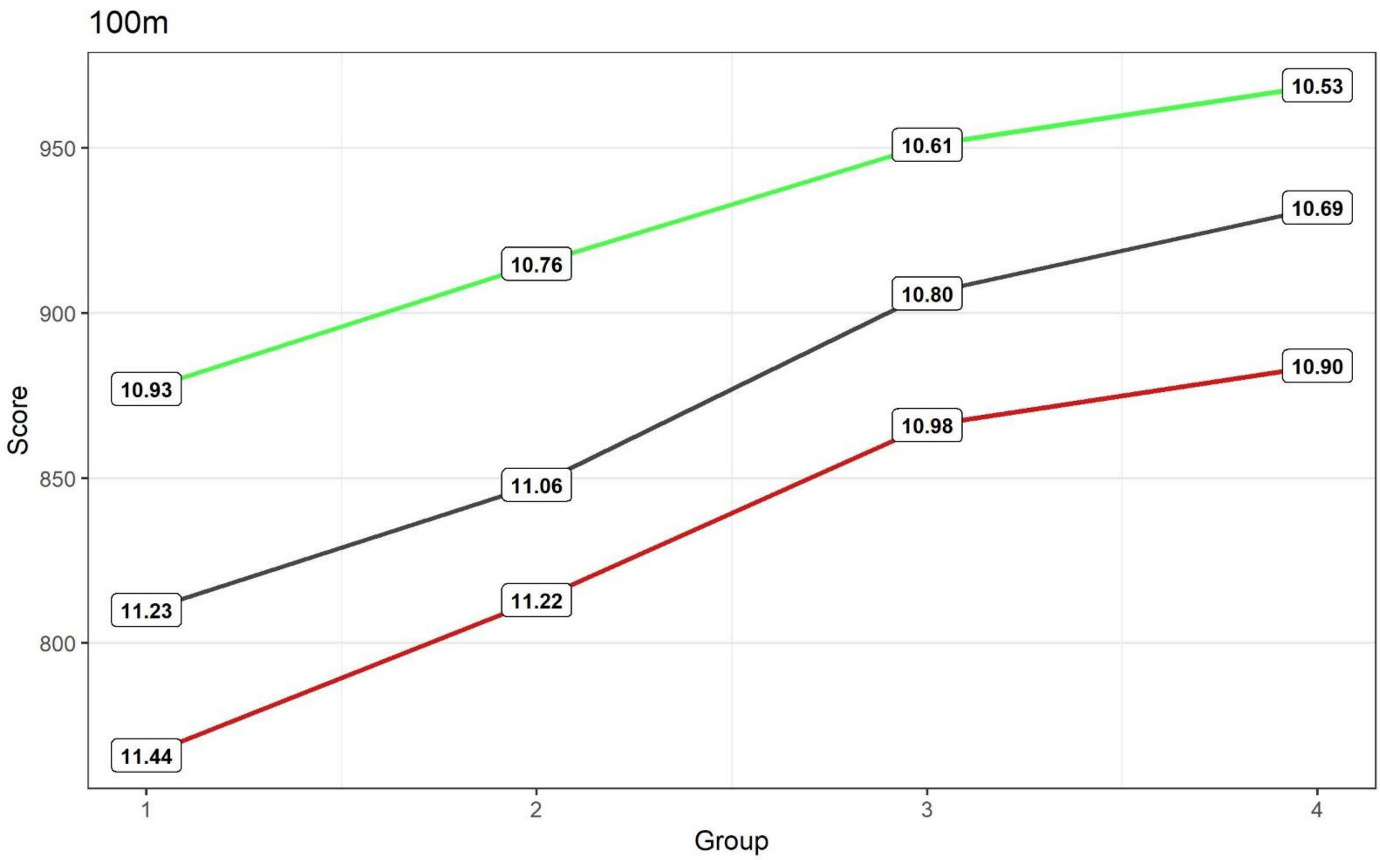
$$\mathbb{E}\left(\boldsymbol{\beta}_i \middle| \boldsymbol{y}_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_R\,, \Gamma^{-1} \to 0\right) = (\boldsymbol{Z}_i^\top\boldsymbol{R}_i^{-1}\boldsymbol{Z}_i + \boldsymbol{D}_i^{-1})^{-1}\boldsymbol{Z}_i^\top\boldsymbol{R}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\alpha}}\left(\widehat{\boldsymbol{\theta}}_R\right)\right)$$

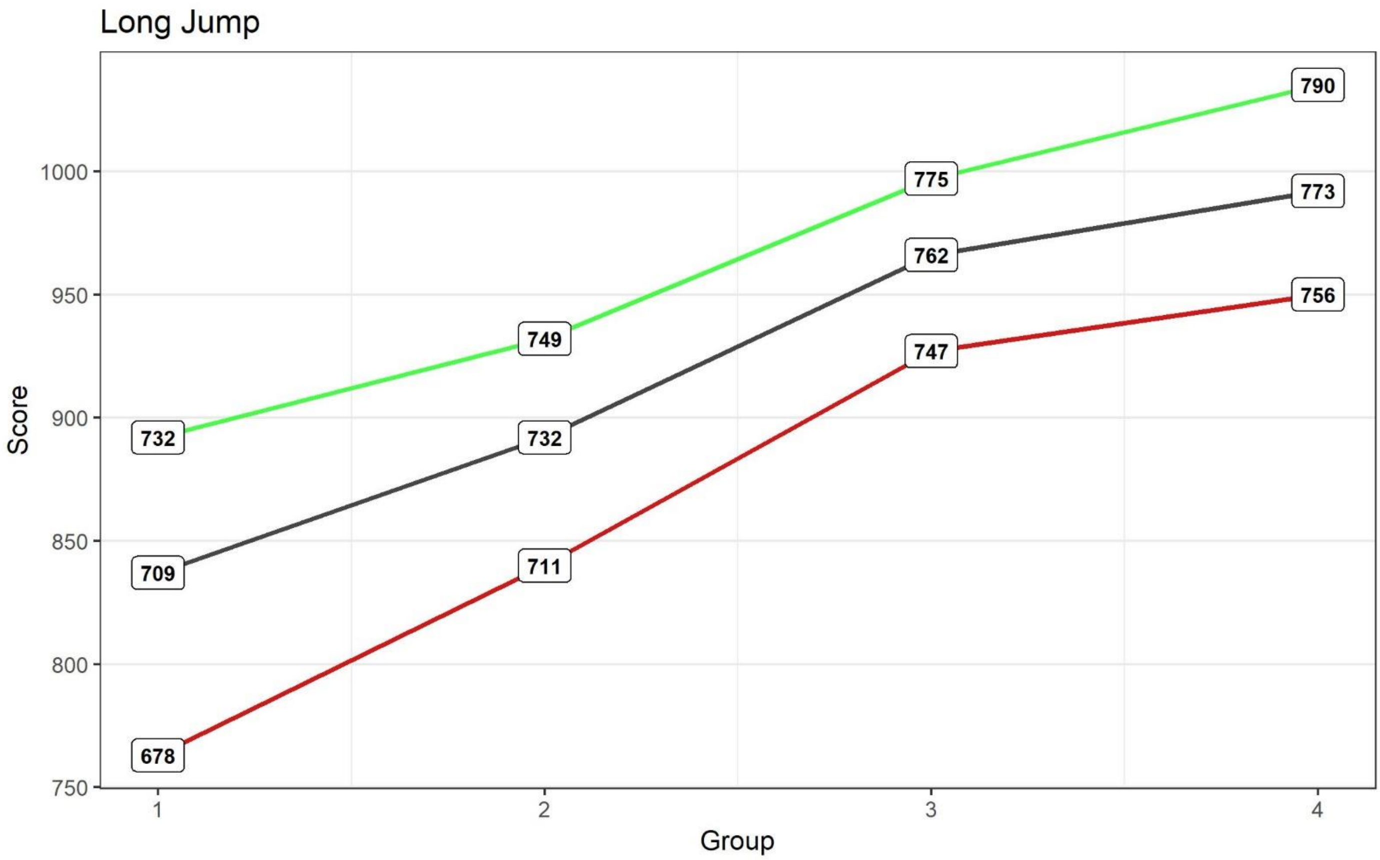
which can be verified as

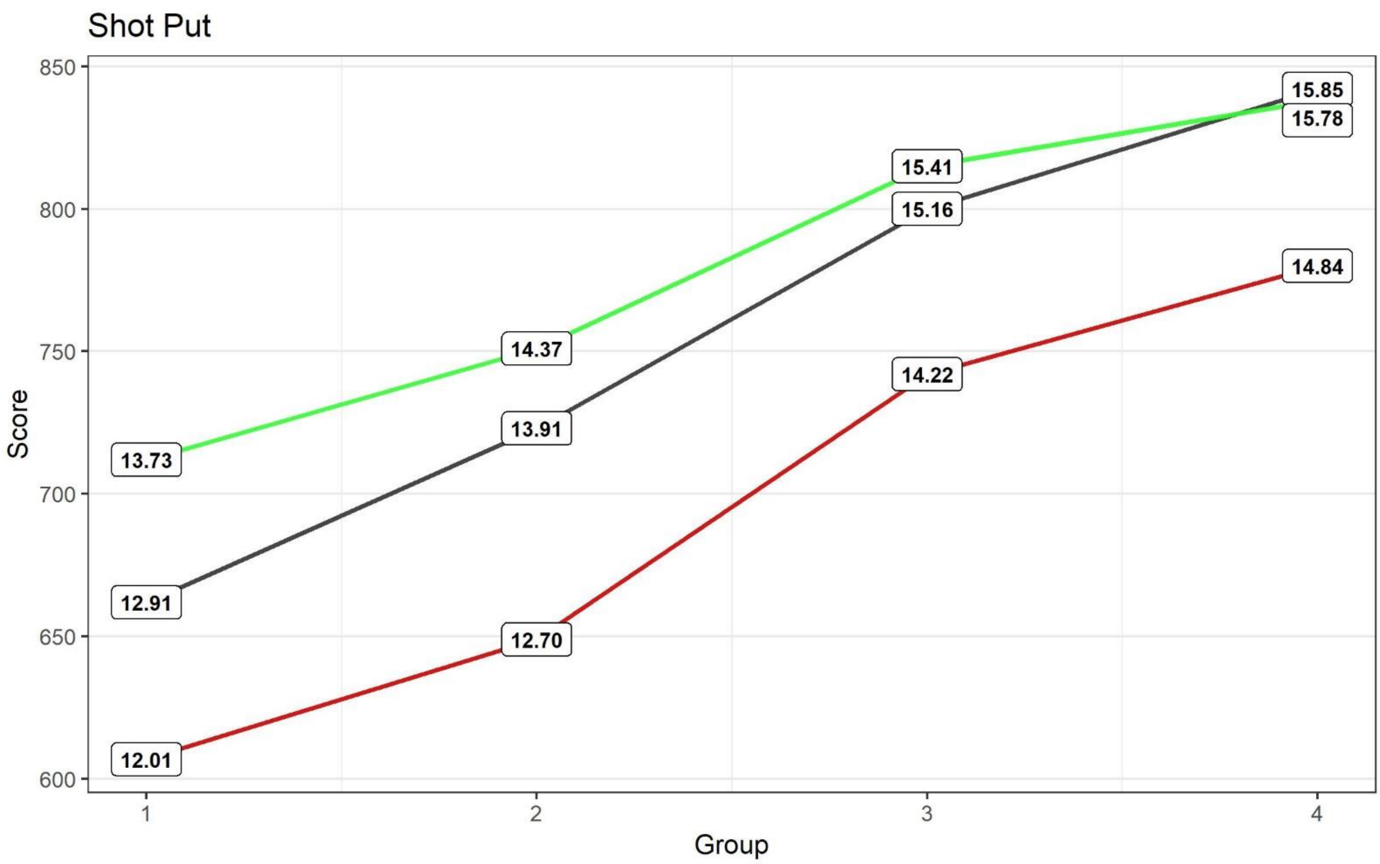
$$\boldsymbol{D}_i \boldsymbol{Z}_i^{\top} (\boldsymbol{R}_i + \boldsymbol{Z}_i \boldsymbol{D}_i \boldsymbol{Z}_i^{\top})^{-1} \left(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\alpha}}\big(\widehat{\boldsymbol{\theta}}_R\big) \right).$$

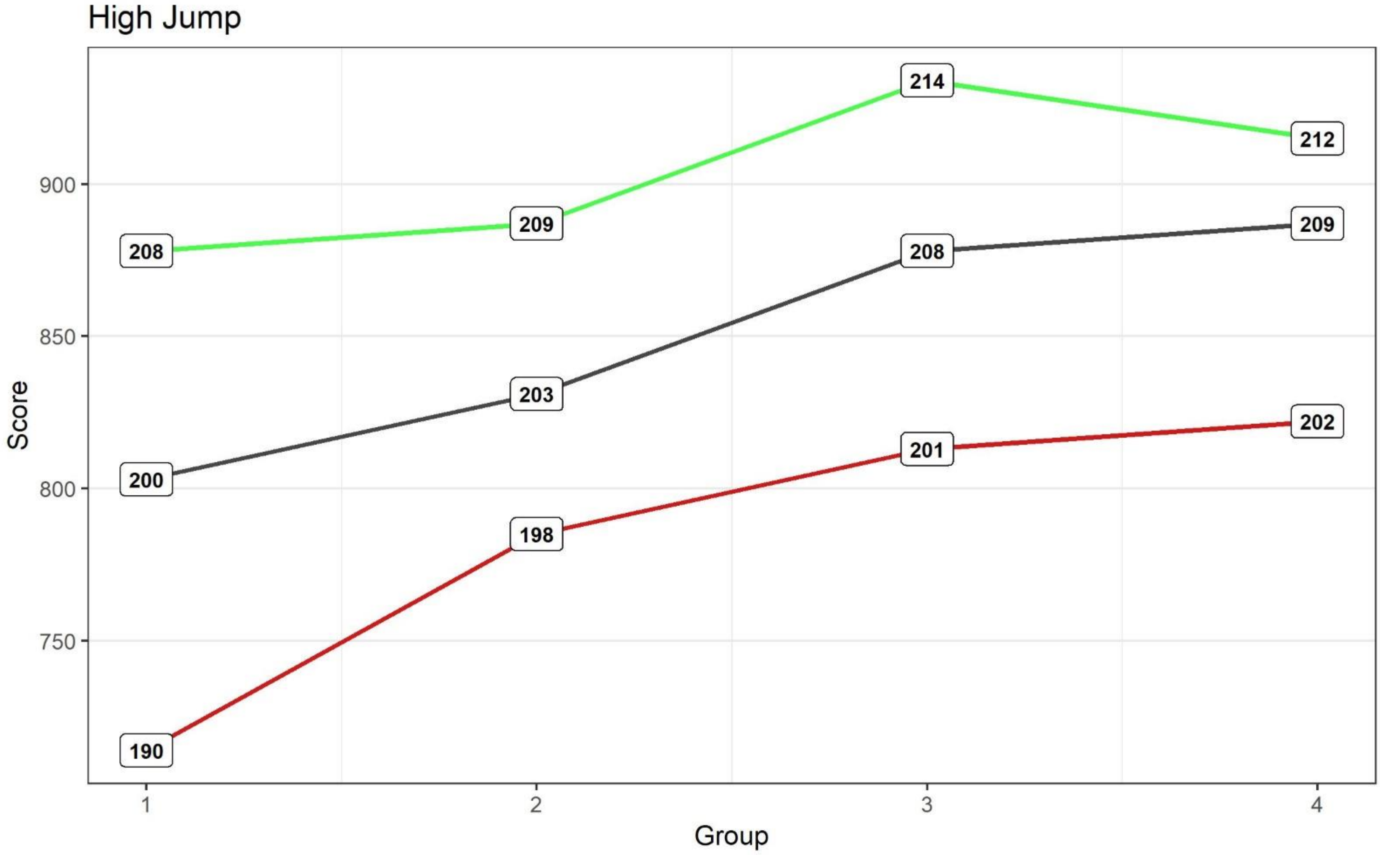
Appendix B

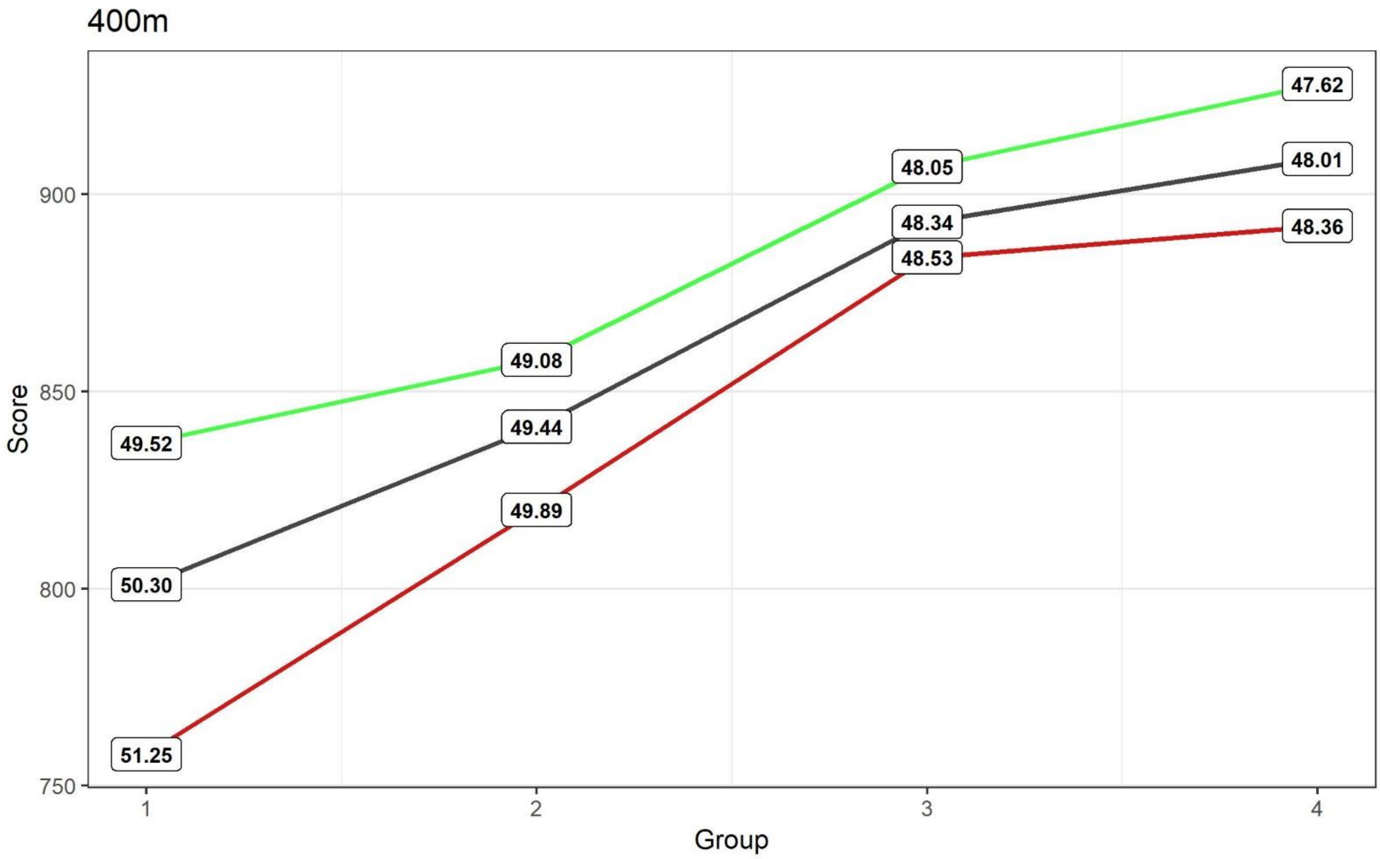
The following figures are showing the median development of three subgroups based on their starting level in each event. Strong performers at level 1 are represented by green and weak performers by red. The rest, here considered "normal" performers, are black. The results matching the points are labeled.

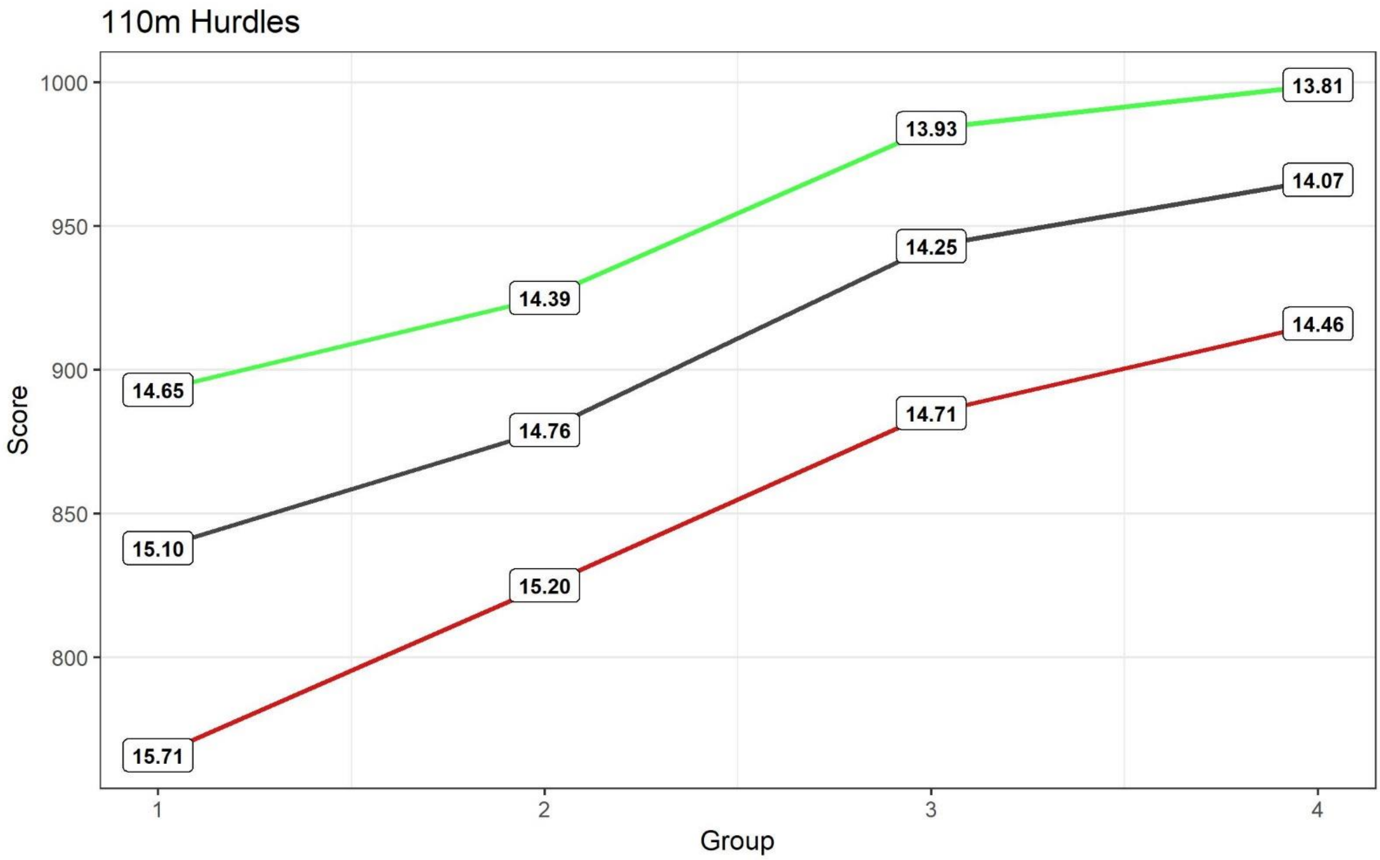


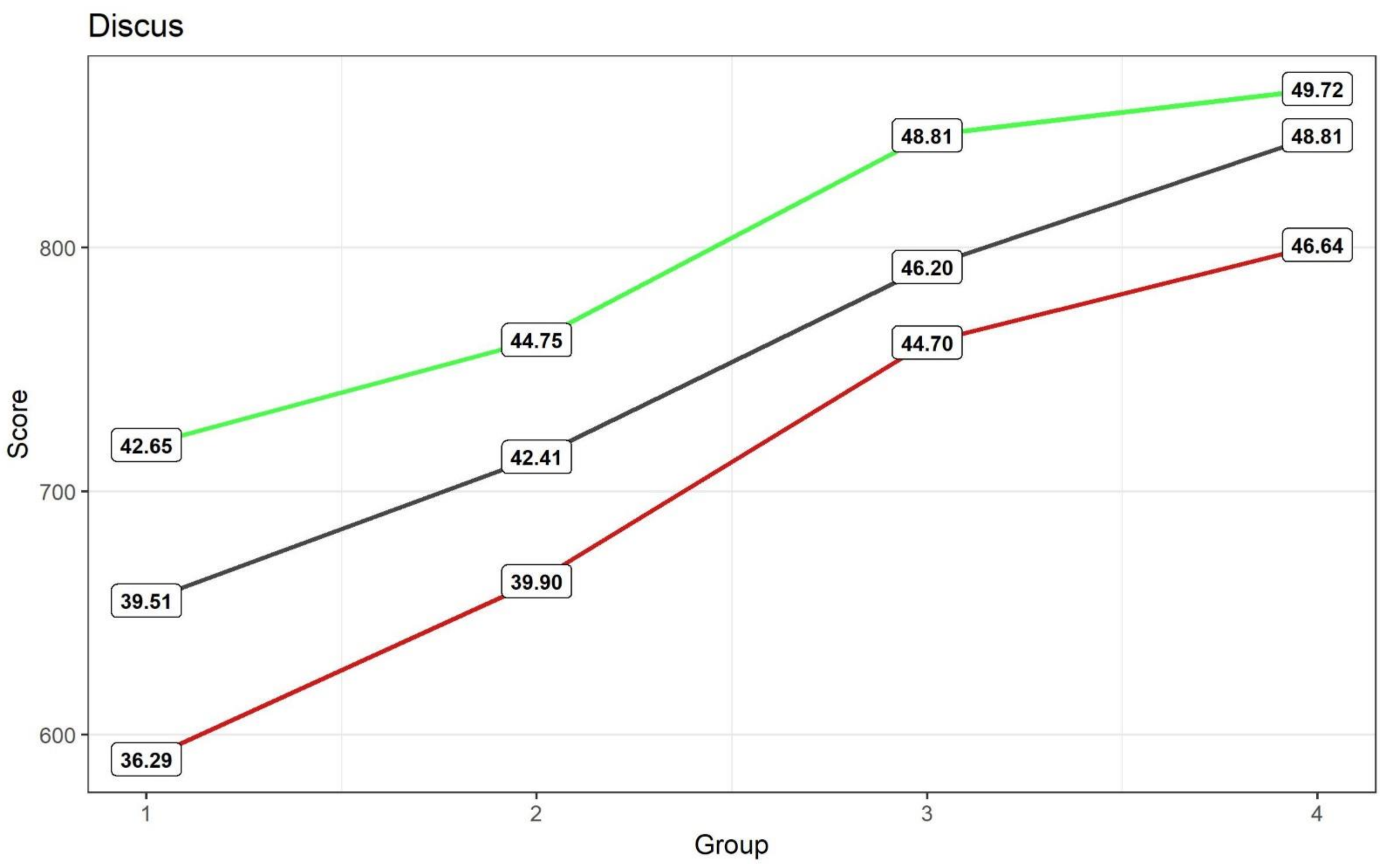

Long Jump
1000
950
900
850
800
750
Score
790
775
773
762
756
749
747
732
732
711
709
678
1
2
3
4
Group

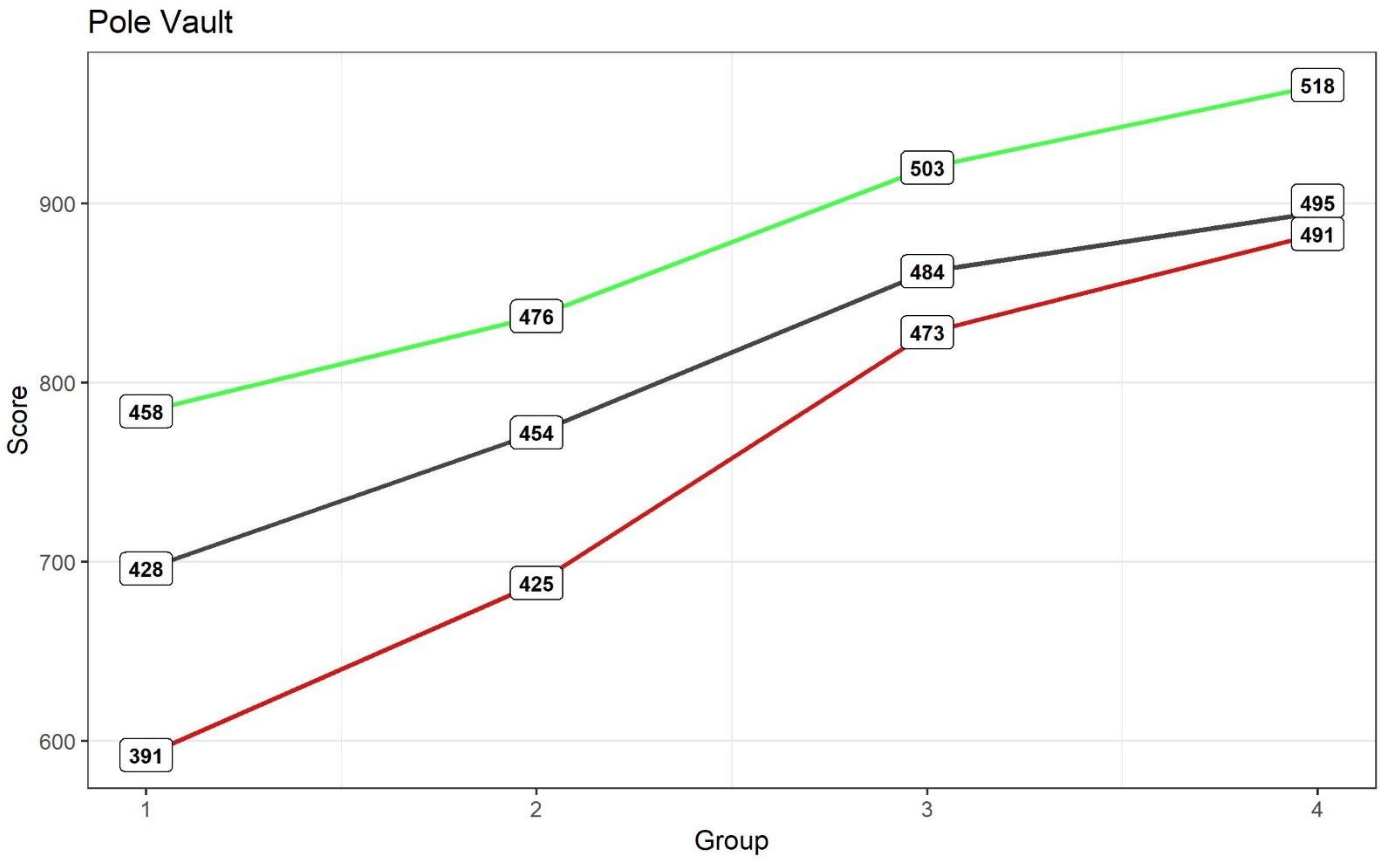

Shot Put
850
800
750
700
650
600
Score
15.85
15.78
15.41
15.16
14.84
14.37
14.22
13.91
13.73
12.91
12.70
12.01
1
2
3
4
Group

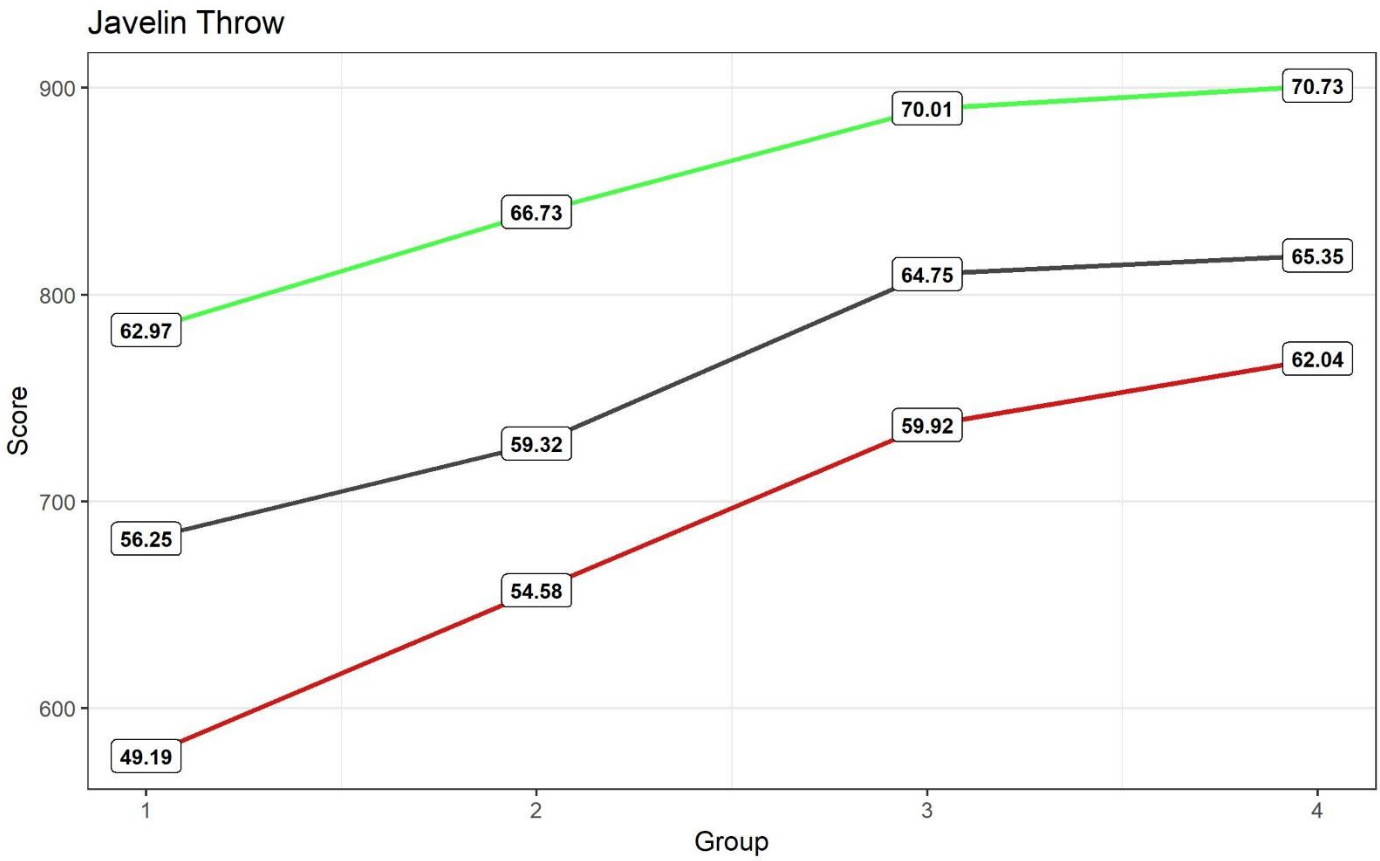

High Jump
214
212
900
209
208
209
208
850
203
Score
202
201
800
200
198
750
190
1
2
3
4
Group

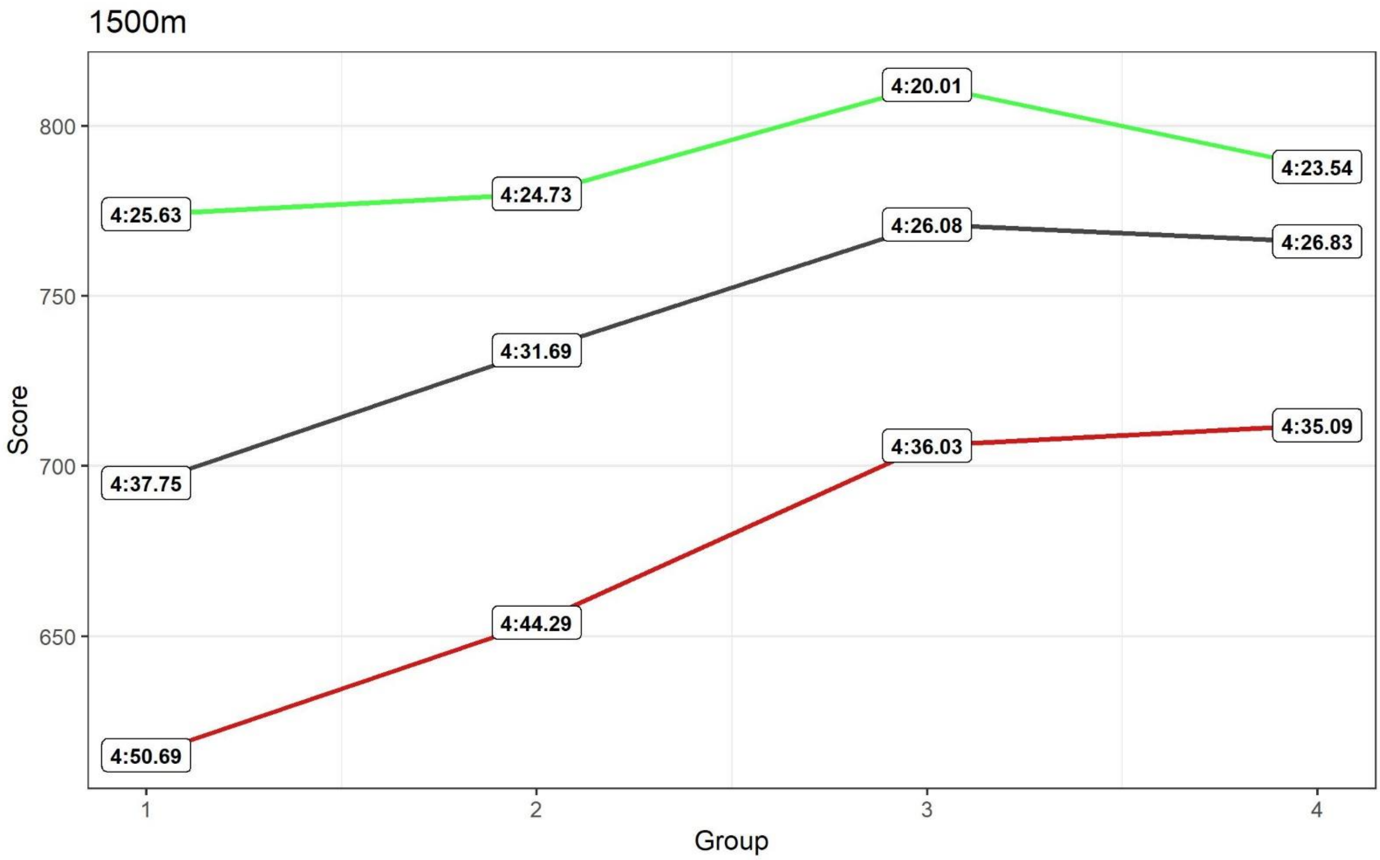

400m
47.62
48.05
48.01
900
48.34
48.36
48.53
49.08
850
49.52
49.44
Score
49.89
800
50.30
51.25
750
1
2
3
4
Group


110m Hurdles
1000
950
900
850
800
Score
1
2
3
4
Group
13.81
13.93
14.07
14.25
14.46
14.39
14.65
14.71
14.76
15.10
15.20
15.71


Discus
800
700
600
Score
1
2
3
4
Group
49.72
48.81
48.81
46.64
46.20
44.75
44.70
42.65
42.41
39.51
39.90
36.29


Pole Vault
Score
Group
900
800
700
600
1
2
3
4
458
476
503
518
428
454
484
495
391
425
473
491


Javelin Throw
Score
Group
900
800
700
600
1
2
3
4
62.97
66.73
70.01
70.73
56.25
59.32
64.75
65.35
49.19
54.58
59.92
62.04


1500m
800
750
700
650
Score
4:25.63
4:24.73
4:20.01
4:23.54
4:26.08
4:26.83
4:31.69
4:37.75
4:35.09
4:36.03
4:44.29
4:50.69
1
2
3
4
Group